Towards the Development of the TPR Scaffold into Novel Biomaterials & Bioswitches

Charlotte Millership

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Statement of Originality

I, Charlotte Millership, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below. I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: 19/09/2014

Details of collaborations and publications:

The Heteropolymer Ising model was developed by Dr. J.J. Phillips (Section 4.3.6). NMR data was collected and raw data was processed by Dr. Toms (Chapter 4). Cloning and expression of azide mutants was performed with Joseph Harvey (Chapter 5).

J. J. Phillips, Y. Javadi, C. Millership, and E. R. G. Main. Modulation of the Multistate folding of Designed TPR Proteins through Intrinsic and Extrinsic Factors. *Protein Science*, 21(3):327-338, 2012.

J. J. Phillips, C. Millership, and E. R. G. Main. Fibrous Nanostructures from the Self-Assembly of Designed Repeat Protein Modules. *Angewandte Chemie*, 51(52): 13132-13135, 2012.

E. R. G. Main, J. J. Phillips, and C. Millership. Repeat Protein Engineering: Creating Functional Nanostructures/Biomaterials from Modular Building Blocks. *Biochem. Soc. Trans.*, 41(5):1152-1158, 2013.

C. Millership, J. J. Phillips, and E. R. G. Main. Creating a Thermodynamically-triggered Conformational Switch in Consensus Repeat Proteins. *(Under submission)*

Abstract

TetratricoPeptide Repeats or TPRs are a class of repeat proteins made up of -helices. Each repeat contains 34 amino acids that form a helix-turn-helix motif and is stabilised by short range interactions creating a non-globular fold. Tandem arrays of these repeats form stable superhelical structures. The modular nature of the TPR fold has allowed a series of consensus TPRs (CTPRs) to be designed where the number of repeat units has been varied.

We have exploited the modular nature of CTPR proteins in order to create fibres via a bottom-up approach. Using Native Chemical Ligation (NCL) we have been able to trigger specific assembly of monomeric CTPR units to form extended fibrous structures up to microns in length (as viewed by TEM). This reaction proceeds at room temperature and neutral pH, with filaments observed within 12 hours.

The equilibrium unfolding of CTPRs is prone to the population of partially folded states. Through studying the stability of a series of deletion mutants and using a Heteropolymer Ising model to analyse the unfolding data we have been able to design a CTPR with a conformational 'switch'. This new CTPR was designed to populate a stable intermediate, with an exposed dimerisation interface, under certain conditions. When this new construct was analysed using 2D NMR and CD spectroscopy, it was found to selectively unfold its C-terminal α-helix at a specific concentration of GuHCl. Our aim is to develop a system in which a 'switching' CTPR is used as a sensor that, when triggered by environmental conditions, partially unfolds and oligomerises.

Acknowledgments

I've had a fantastic 4 years at Queen Mary and first of all I'd like to thank the all the lovely people in 4.36 for all their help and support, in no particular order they are: Emma, Anna, Amy, Jun, Erica, Petra, Max, Helen, Anya, Tatiana, Sam, Andy and Ricardo. Special thanks to Emma for sending me with cute cat pictures when I needed a break. Thanks to the members of 'teagroup', for making me laugh during my morning intake of caffeine. Big thanks to June for providing me with tea and lemon cake, without which, I probably wouldn't have done any work.

Thanks to the members of lab G35, for letting me use their lab for various experiments and eat all their birthday cakes. Especially I'd like to thank Nadine and Helen for all their help with using the CD machine, both of you (along with Christian and Joe) made my experiments in G35 a lot of fun. Thankyou to Prof Pickersgill and Dr Viles for providing scientific insight and support during the course of my PhD.

I'd like to thank Dr Luklinska and Graham McPhail and Dr Nield for their help with using the TEM. Thanks to Dr Harold Toms for collecting my NMR data.

Thanks to my fellow PhD students in the Main group. Thanks to Yalda for encouraging me to do a PhD. Thanks to Joe for being so tidy in the lab and for being such a good student. Thanks to Sunny for always helping me to fix broken equipment and always taunting me with your awesome lunches (ding!). Thanks also to James for buying more printer ink just in time for me to print my thesis.

A big thanks to my dearest friends JJ and Ru. Thanks to JJ for being so patient with me when I was just a clueless undergrad that had never been in a biochemistry lab. Thanks to Ru for your help, both scientific and personal, thanks for always being there for a chat and a tea break. Mostly I want to thank my supervisor, Ewan, for all his encouragement, help and support and for believing in me.

Thanks to my wonderful partner, Chris Perkins, for supporting me throughout my 8 years at university.

For my partner Chris Perkins

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Proteins

Proteins are one of the four biopolymers that are essential to life. They perform a crucial role in most biological processes. This ranges from catalysing reactions within cells to forming the structure of the cells themselves [1, 2]. The ability to understand the important role that proteins play in biological processes was vastly increased when X-ray crystallography made it possible to see their atomic level structure. This allowed the relationship between the structure of proteins and their function to be better understood. For example by defining the active sites of an enzyme we can better understand the mechanism of catalysis [3, 4].

Since the first reported structures fifty years ago, there are now nearly a hundred thousand structures, solved through X-ray crystallography and NMR analysis. These have revealed that there are a limited set of common motifs. The wealth of structural data has led to the classification of proteins by specific folds. Protein structures are deposited to the RCSB protein database, where they are arranged into categories according to their origin and function. Other databases, like SCOP (Structural Classification of Proteins) have grouped proteins together that share

common structural characteristics [5, 6]. In doing this, it is possible to construct common sequences that tend to define certain folds. By classifying proteins in this manner it is possible to characterise how structure and to some extent 'function' are related to the primary sequence of the protein. Comparison of amino acid sequences using such databases have shown that sequences that are highly homologous form similar motifs and domains. It is now possible to take a sequence of a proteins with an unknown structure and search the structural databases for proteins with similar sequences and, if a match is found, the fold can be estimated.

1.1.1 Classification of Protein Structure

We can divide proteins into 2 main categories: (i) globular (the machinery of cells) and (ii) fibrous (the structural components).

(i) Globular proteins typically contain a mix of different motifs, where the folding is driven by the need to create a hydrophobic core and a hydrophilic surface on the outside of the protein [1, 7]. The free energy between the folded and the unfolded states of globular proteins is usually between 5-15 $kcalmol^{-1}$. This is quite small considering that the energy from a single hydrogen bond is 2-5 $kcalmol^{-1}$ [7]. The large variation in structures mean that globular proteins can provide a large variety of functions, such as catalysis, molecular recognitions and cellular transport.

(ii) Fibrous proteins, in contrast to globular, form elongated repetitive structural elements and typically serve as structural materials within cells because of their strength. An example is collagen, which is comprised of 3 chains a thousand residues long that coil around each other [1]. These chains then form larger fibrous structures that give cell walls strength [7].

Interestingly there are some protein scaffolds that seem to bridge both of these def-

initions, such as linear repeat proteins. These 'novel' folds (discovered in the 1990s) possess a hydrophobic core like globular proteins but are repetitive in sequence and structure like fibrous proteins.

1.1.2 Linear Repeat Proteins

1.1.2.1 Structure of Repeat Proteins

Linear repeat proteins are characterised as tandem arrays of short structural motifs that form a single protein chain [8–10]. Repeat motifs are between a 10-40 amino acids long and can be repeated from as little as twice to over 20 times. The arrayed structural motifs form either a domain of a protein or simply the whole protein. Repeat proteins are typically classified according to their tertiary structure [10]. Repeat motifs can be all α-helical, β-helical or a combination of both. For example, Figure 1.1 shows the crystal structures of an α-helical Tetratricopeptide protein (B), a β-helix repeat (D) and a α-helix and β-strand-containing Leucine-rich protein (F) with each repeat coloured differently.

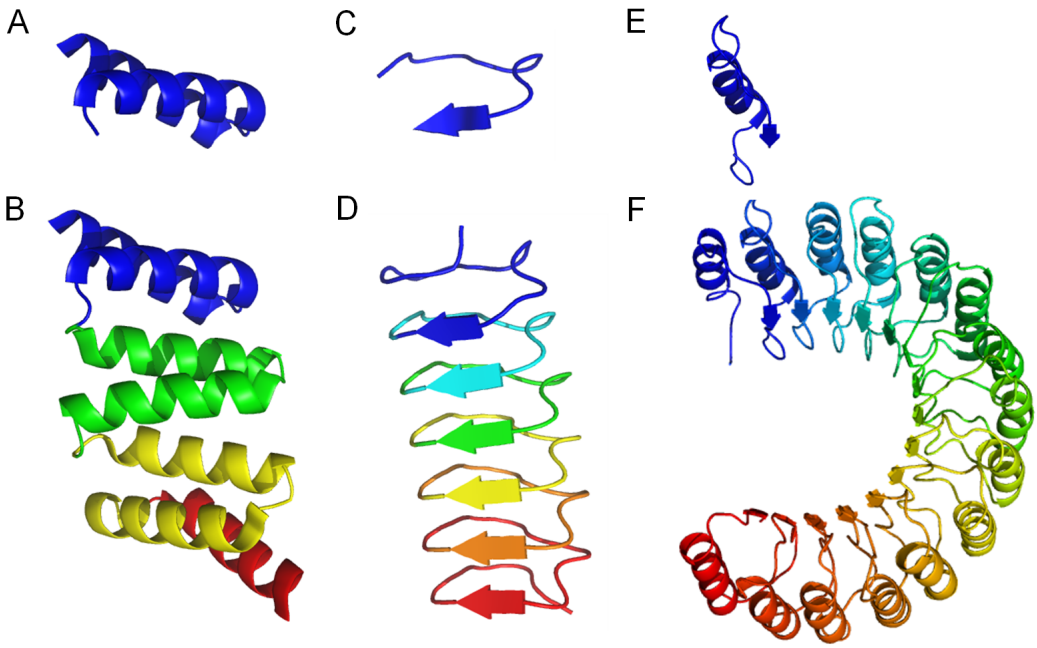

Figure 1.1: Repeat motifs can be all α-helical, β-helical or a combination of both. Structure of a single repeat (A) from a consensus Tetratricopeptide (1NA0) with each repeat shown in a different colour (B). Structure of a single repeat (C) from a β-helix repeat (1EZG) with each repeat shown in a different colour (D). Structure of a single repeat (E) from a Leucine-rich repeat (1A4Y) with each repeat shown in a different colour (F).

The principal feature of linear repeat proteins is that the final 3D fold is formed and stabilised only through interactions between residues that are close in primary sequence, Figure 1.2. Thus the N and the C-terminus rarely interact. In contrast, the final folded structure of globular proteins are stabilised by local contacts but also from interactions between amino acids that are distant in primary sequence [11], Figure 1.2. This means that globular proteins can be thought of as more 'complex' structures. As repeat proteins are formed through sequence-local contacts, they are tolerant to the addition or subtraction of whole repeat units [8, 11–13]. It is this modularity makes repeat proteins distinctive and different from globular proteins [10].

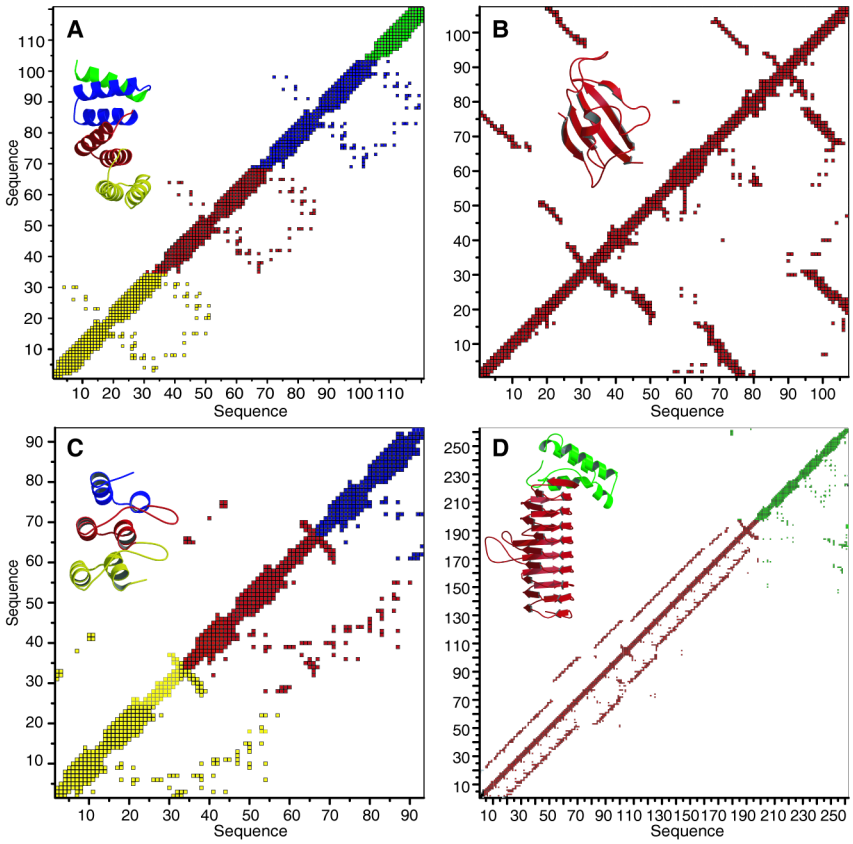

Figure 1.2: Contact order maps for (A) CTPR3 (1NA0), (B) a globular protein, FKBP12 (2PPN), (C) an Ankyrin repeat protein (1N0Q) and (D) and a β-helix protein (1LXA). These diagrams plot a point for every interaction between 2 residues, using their position in the primary sequence as the co-ordinates. These diagrams elegantly show how in globular proteins, interactions occur between residues that are distant in primary sequence. Where as in repeat proteins the contacts are formed between residues that are close in primary sequence. Diagram from reference [13]

1.1.2.2 Functions of Repeat Proteins

Repeat proteins are important for many biological processes such as cell adhesion [14], signal transduction [15] and disease resistance [16–18]. They perform this wide range of roles by mediating protein-protein interactions [13, 19]. This exciting binding repertoire of proteins and peptides is enabled through hyper-variable positions within their sequence that can be varied without affecting the fold [13, 19]. In the next section we will describe a representative selection of repeat proteins and their functions, shown in Figure 1.3.

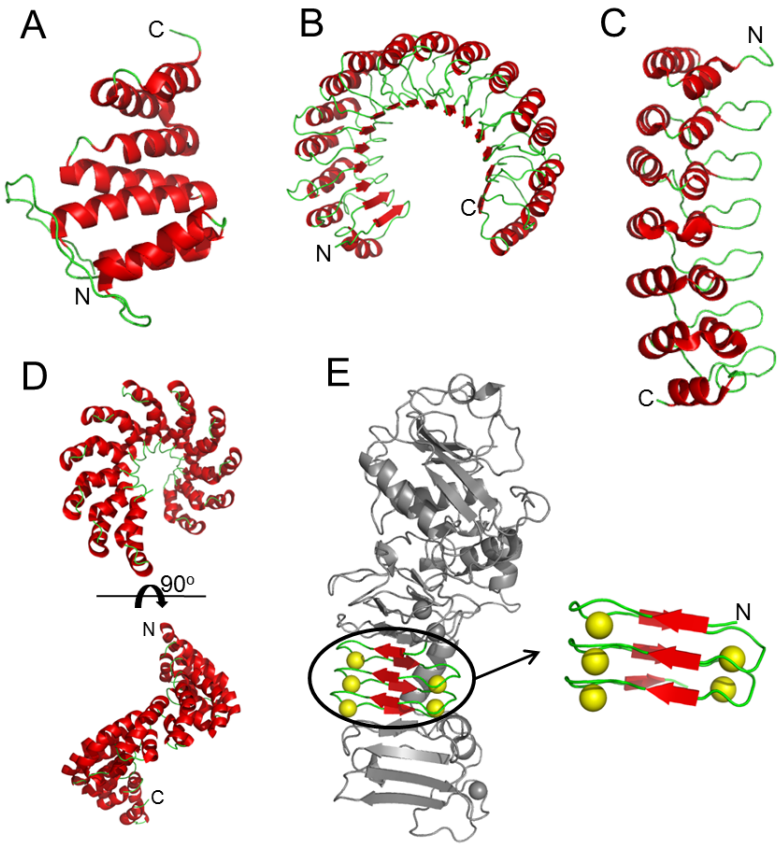

Figure 1.3: (C) Crystal structure of the Ankyrin repeat domain of Bcl-3 (1K1A). (B) Crystal structure of a Leucine-rich repeat protein that is a human placental RNase inhibitor (1A4Y). (C) Crystal structure of the HEAT repeat protein from PR65A (1B3U). (D) The solution structure of a TPR of human smooth muscle cell associated protein-1, isoform 2 (2DBA). (E) Crystal structure of a TAL effector DNA-binding domain of dHax3 (3V6P) in 2 orientations. (F) Crystal structure of the alkaline protease of Pseudomonas aeruginosa, that contains a parallel β-roll motif (1KAP), that is bound to Ca^{2+}.

1.1.2.2.1 Ankyrin

The Ankyrin class of repeat proteins were first discovered in 1987 [20] and are named after the human erythrocyte protein Ankyrin [10]. A single repeat consists of 33 amino acids, which form 2 α-helices and a β-hairpin. Ankyrin repeats bind to protein surfaces through a hyper-variable loop region. For example, 53BP2, contains a domain 4 Ankyrin repeats and binds to the tumour suppressor p53 through residues on the β-hairpin loop between the 3rd and 4th repeats [21], Figure 1.4. Thus binding motifs are located on the loop region, on the concave face of the protein.

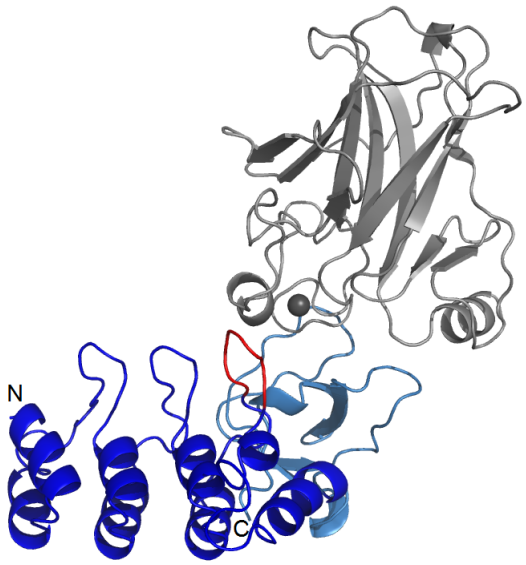

Figure 1.4: The protein 53BP2 in complex with the p53 tumour suppressor (grey) (1YCS). The 53BP2 contains a 4 repeat Ankyrin domain (blue), the loop that binds to the p53 protein is highlighted in red.

1.1.2.2.2 Leucine-Rich Repeat

The first crystal structure of the Leucine-rich repeat proteins was reported in 1993. The repeat is defined by a conserved sequence of leucine (or valine or isoleucine) residues either 11 or 12 in length, with the general formula: LxxLxLxxNxL or LxxLxLxxCxxL, n= asparagine, threonine, cysteine or serine and C= cysteine or serine [22]. A single repeat is typically between 20-30 amino acids long and they form a β-strand-turn-α-helix motif [10, 22]. As the repeats stack together they form an arch, where the binding site is usually located [10]. LRR proteins are involved in a wide variety of functions such as disease resistance in plants, enzymatic activity and cell adhesion [23]. An example of a LRR protein is the GTPase-activating protein, rna1p. This protein contains 11 repeats, in the loop region after the 3rd repeat there is an insertion of the motif GRLxxE. This solvent exposed loop contains the conserved residues arginine and glutamate are required for binding to the Ran protein and the former is required for the catalytic activity [24], Figure 1.5.

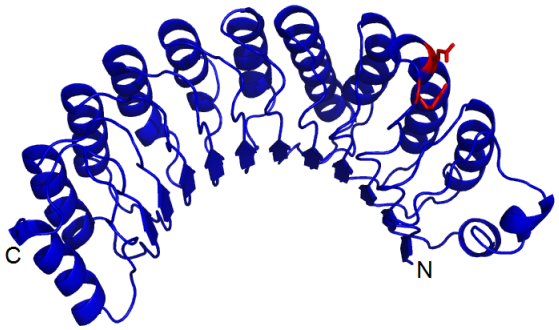

Figure 1.5: LRR protein is the GTPase-activating protein, rna1p (1YRG)

1.1.2.2.3 TetraTricopeptide Repeat

The TetraTricopeptide Repeat (TPR) represents a large family of linear repeat proteins. Identified in 1990 the TPR fold is ubiquitous and is known to be involved in a variety of functions including protein transport and protein folding [9, 25]. Each repeat contains a minimum of 34 amino acids, which form a helix-turn-helix motif. The two α-helices, often referred to as the A and B helix. In naturally occurring proteins this motif is repeated from 3 to 16 times, with each repeat stacking on top of the other to form an extended structure that is superhelical [13]. TPR proteins often bind peptides. For example the Hop protein contains 3 TPR domains that bind to the C-terminal regions of Hsp70 and Hsp90. The binding pocket is located on the concave face of the TPR and forms electrostatic interactions with the motif EEVD [26], Figure 1.6.

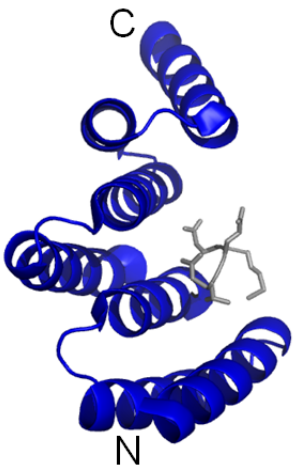

Figure 1.6: Crystal structure of the TPR domain of Hop in complex with the C-terminal peptide sequence of Hsp90 (1ELR)

1.1.2.2.4 HEAT/Armadillo repeat

Armadillo [27] and HEAT [28] repeats were first identified in 1989 and 1995 respectively. Structurally they share the same C-terminal helix, but differ in the rest of the structure. Armadillo repeats contain 3 α-helices and HEAT repeats contain 2 [10]. These repeats bind to proteins via interactions on the concave face. An array of HEAT repeats form the importin-β that binds to the IBB domain of importin-α. The inner surface of the super helix that the HEAT repeat creates the binding pocket for the IBB domain. The IBB binds through contacts between its basic residues and the acidic residues on the inner helix of repeats 12-19, and also through the acidic loop on repeat 8 [29].

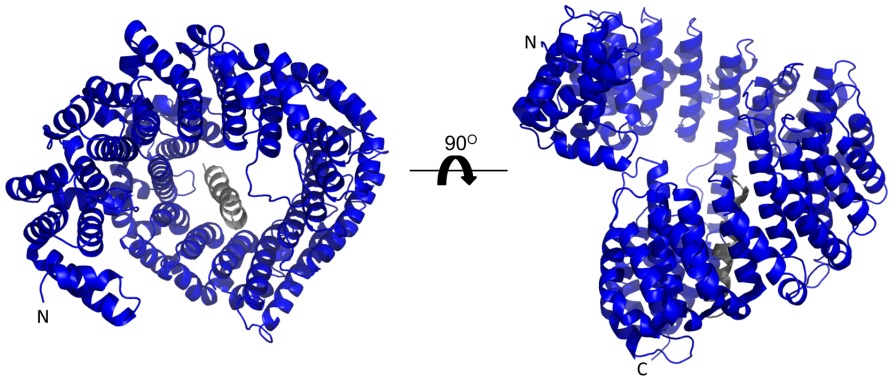

Figure 1.7: The HEAT repeat protein, importin-β, bound to the IBB domain (grey) of importin-α (grey) (1QGR).

1.1.2.2.5 Transcription Activator–Like

Transcription activator–like (TAL) effectors were first identified in 1989 and are secreted by plant bacteria [30]. Once secreted into cells, the TAL binds to specific DNA promoters leading to protein expression [31]. The TAL repeats are characterised by an α-helix-turn-α-helix motif that is 34 amino acids long. The repeats stack to form a superhelical structure that binds to DNA [32]. The interesting feature of the TAL repeats is that the sequence is highly conserved between repeats [30]. Hyper-variable residues are found at positions 12 and 13, located in the loop region between the α-helices and on the inside face of the superhelical structure. The residue at position 13 binds to the DNA base and residue 12 stabilises the conformation of the DNA binding residue [33]. When bound around DNA, the superhelical structure contracts into a helix with a shorter pitch [30], Figure 1.8. Each TAL repeat binds specific bases thus the TAL proteins can define a binding site to any specific DNA sequence. TALs have been designed to bind to DNA sequences in order to suppress or activate transcription of specific genes [33].

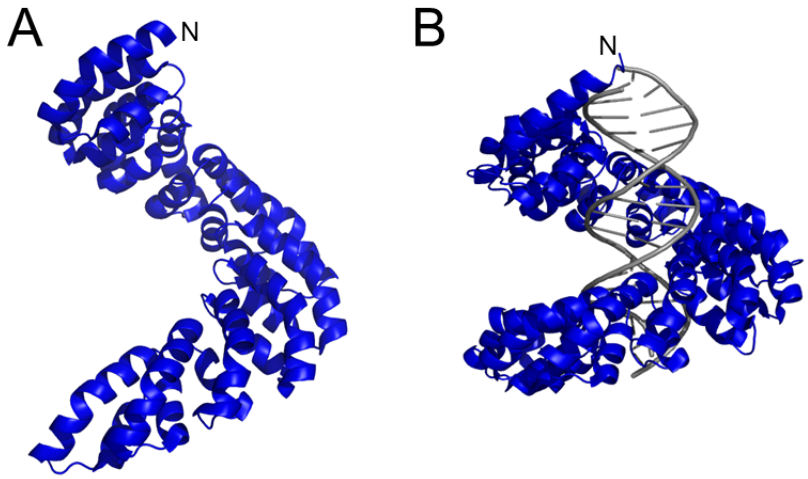

Figure 1.8: A Transcription activator–like effector unbound (A) (3V6P) or bound (B) (3V6T) to DNA. There is a dramatic decrease in the pitch of the helix when the TAL is bound DAN.

1.1.2.2.6 β-helix and β-roll motif

The β-helix and β-roll motif were first discovered in 1993. These repeat proteins consist of β-strands that align to form parallel β-sheets, along the length of the helix [34, 35]. They exist in a diverse range of lengths from the small antifreeze proteins, with only 12 residues per repeat to the 18 residue repeats of the β-roll motifs. The β-roll motifs are stabilised by binding Ca^{2+} atoms on the loop region [36]. The β-helix from the *Tenebrio Molitor* is used to non-colligatively suppress the freezing point of water. Each turn of the helix has a single β-strand that has a TCT motif, the threonine residues are solvent exposed and positioned to bind to the face of an ice crystal [37], Figure 1.9.

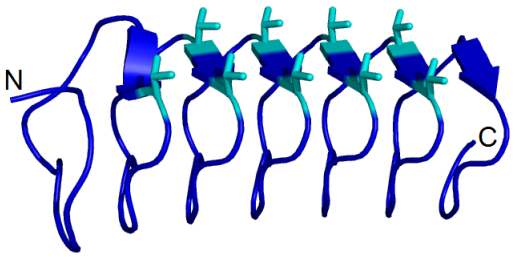

Figure 1.9: A β-helix antifreeze repeat (1EZG), the threonine residues along the β-strands bind to the surface water molecules of ice.

1.1.3 Designing Repeat Proteins

As can be seen from the previous section, it is clear that repeat proteins have an extremely flexible structure, i.e. they can be increased or decreased in size through addition or subtraction of whole repeat units and form the most amazing protein-protein interaction motifs. Significantly, the residues that are important for repeat binding are easily delineated from those important for structure. This makes them similar to antibodies and consequently they are exciting design targets which have generated great interest.

1.1.3.1 Consensus Design

The first successfully designed repeat proteins were achieved through a consensus design approach. Consensus design involves collating all the available sequences of proteins that adopt a particular repeat fold. The sequences are divided into repeat motif and a multiple sequence alignment is performed to produce a consensus sequence that defines the most common residues at each position in the motif. The designed repeat can then be repeated within a gene to produce a protein with identical motifs. This technique was first successfully used to design the Ankyrin [38] and TPR [8] proteins. Figure 1.10A and B shows the consensus sequence for an Ankyrin and a CTPR from references [39] and [8] respectively. Both proteins contain 2 α-helices per repeat and the Ankyrin repeat also has a -hairpin, Figure 1.10E and F.

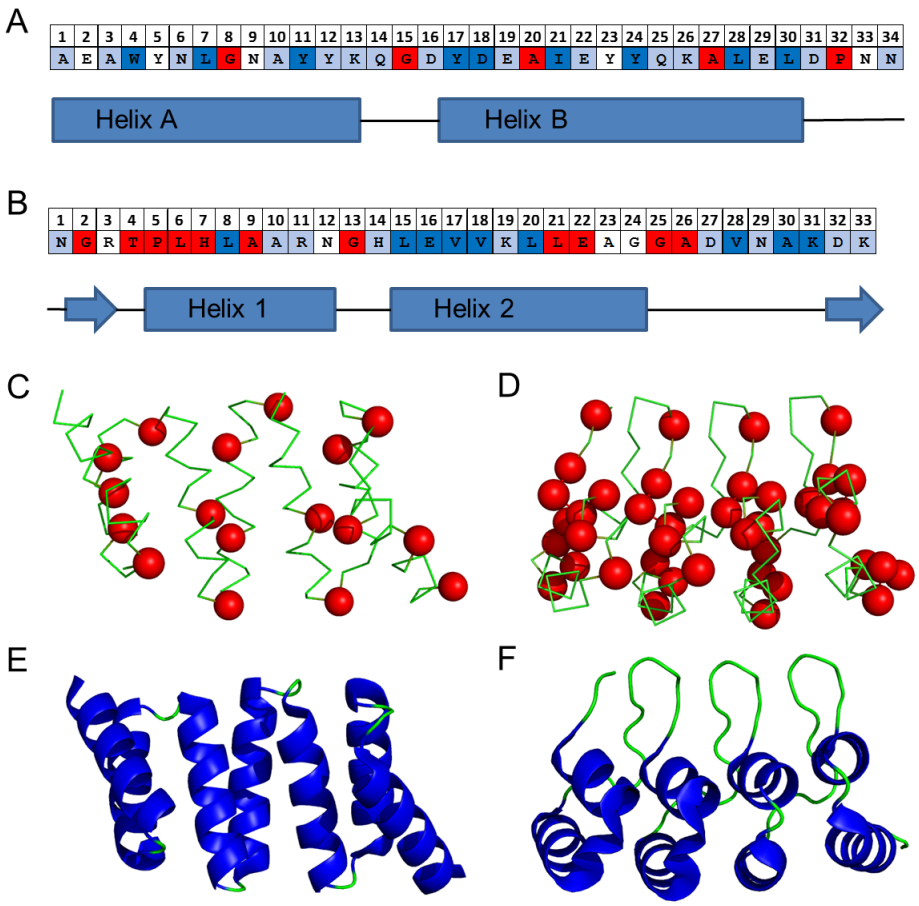

Figure 1.10: (A) Sequence of a single CTPR repeat, that contains 2 α-helices separated by a short loop. The highly conserved residues are highlighted in red. Semi-conserved residues are highlighted in blue in the consensus sequence. These residues are divided into two types where either the high frequency residues for that position share the same properties (dark blue) or have different properties (pale blue) [13]. (B) Sequence of a single designed Ankyrin repeat, that contains 2 α-helices and a -hairpin. The highly conserved residues are highlighted in red. Semi-conserved residues are highlighted in blue in the consensus sequence. These residues are divided into two types where either the high frequency residues for that position share the same properties (dark blue) or have different properties (pale blue) [39]. (C)The conserved structure-forming residues mapped onto the structure of CTPR3 (1NA0). (D)The conserved structure-forming residues mapped onto the structure of ANK4 (1N0R). (E) Structure of CTPR3 (1NA0). (F) Structure of ANK4 (1N0R).

Consensus designed proteins were found to be much more stable than their naturally occurring counterparts [8,39]. This is because consensus design selects residues that are structure forming. However, a consensus design approach removes the functional binding motifs, to leave a bare structural scaffold. Since the initial design studies, there has been development of the consensus repeat protein scaffold into functional

proteins, through re-engineering of binding surfaces.

1.1.3.2 Re-engineering in Binding Motifs

The two most studied systems (Ankyrins and TPRs) were also the first designed proteins to have binding motifs re-engineered onto the bare scaffold. A binding site for the C-terminal section of the Hsp90 protein was grafted onto a CTPR scaffold through a rational design approach. This method involved comparing the sequence for the Hsp90 binding TPR with that of CTPR3, and identifying the residues that are important for binding to the sequence MEEVD. These residues were then grafted onto the CTPR scaffold to give protein, CTPR390. This protein was able to bind to the pentapeptide with a K_d of 200 μM [40]. The structure of CTPR390 is nearly indistinguishable from that of CTPR3 and the thermostability is also similar, this demonstrates that the addition of the binding motif does not require a compromise to the structure or the stability [40].

In contrast, designed Ankyrin repeats (DARPins) have been engineered to bind to specific protein targets through the directed-evolution technique: ribosome display. This technique involves creating a library of DARPin genes, where 7 residues in each repeat are randomised, these are expressed by the ribosome in vitro [41].

Successful binders are selected by 'panning', this involves mounting the binding target to a solid support and incubating with the DARPin library. Once washed with the buffer those proteins that aren't successful will not bind and can be discarded [41]. Those which bind can be eluted and amplified and the binding process repeated. In this way strong binding proteins are evolved. DARPins were initially selected and shown to bind to the human epidermal growth factor receptor 2 (Her2) [42]. Since the platform technology has been commercialised by the company Molecular Partners. Now the technology is being further used as antibody

replacements and are being developed for the treatment of a variety of diseases including retinal diseases such as wet age-related macular degeneration. DARPins have been designed to bind many targets for example: either the inactive (non-phosphorylated) or the active (phosphorylated) form of ERK2, a mitogen activated protein kinase [43]. In another study DARPins with an extended loop region able to bind specific targets was developed. These LoopDARPins were successful in binding ERK2 and the BCL-2 family of proteins, which are essential for regulating apoptosis [44].

The ability to introduce functionality into designed repeat proteins, either through directed evolution or rational design makes them attractive as systems to engineer and design. Moreover, their high stability and modular architectures make them suitable as potential building blocks for biomaterials.

1.2 Biomaterials

Biomaterial is a broad term that encompasses any material that is composed of biological components. For example, spider silk is a biomaterial made of polypeptide fibres containing repeating sequences of non-polar and hydrophobic amino acids [45]. The term can also be used to describe biocompatible materials such as titanium-polyethylene joint replacements. The focus of this thesis is biomaterials made from proteins, therefore we will only discuss biomaterials made from polypeptides.

Engineered biomaterials comprised of peptides and proteins can be ordered into 3 categories [46]. (1) Recombinant proteins designed from naturally occurring biomaterial-forming proteins, such as silk [45] and elastin [47]. (2) Peptide-polymer materials that are composed of short peptides with a polymer attached at one end, for example peptide amphiphiles [48]. (3) Self-assembling peptide systems, such as

coiled coils [46].

There are many potential uses of biomaterials, particularly in the area of drug delivery [49] and tissue engineering [50]. Self-assembling coiled coils have been engineered to form nanocages that have the potential to be used as delivery vessels for therapeutic drugs [49]. Engineered Elastin-based proteins have been used to form self-assembling hydrogels [51] and films [52] and have also been used for tissue engineering [50]. Biomaterials can also be engineered to change their structure on demand, such as stimuli-responsive peptide hydrogels [53,54]. These systems involve hydrogels that react to specific environmental stimuli such as changes in pH or temperature. These materials have the potential to be used as drug delivery systems for example, a smart gel that releases a drug only when placed in contact with the skin.

Repeat proteins have scope to be extremely useful as building blocks in biomaterials as their modular structures can be stacked to create extended arrays. In addition, they can be functionalised through the engineering of binding sites. For example DARPins have been developed as antibodies [43] and CTPRs have been engineered to posses peptide binding sites [40]. In an interesting study, the peptide-binding ability of CTPRs was used as a way to create hydrogels [55]. CTPR proteins containing 18 repeats with 3 peptide binding sites reversibly form hydrogels when mixed with a PEG-peptide crosslinker. These CTPR hydrogels can be dissolved by increasing the salt concentration demonstrating responsiveness to environmental stimuli. The hydrogel complexes were also able to encapsulate a fluorescent protein and a smaller 'drug-mimetic' fluorescent marker, which was released by increasing the ionic strength. This demonstrated the potential of the CTPR hydrogels to be used for protein or drug delivery [55]. The CTPR hydrogel system uses proteins that were expressed recombinantly, which limits the length of protein. In Chapter 3 we will discuss the successful self-assembly of CTPR proteins into fibres up to microns

in length. A fibre that is 1 μm in length contains approximately 100 repeat units. This method greatly expands the repertoire of structures that can be used in the assembly of biomaterials. For example the fibres could be used to make hydrogels by using bi-functional peptide crosslinkers. The use of longer CTPR filaments could create hydrogels with different properties to those presented by Grove *et al* [55]. In addition, fibres could be decorated using metal-binding peptides providing a synthesis route to creating nano-wires.

Biomaterials can be produced through the bottom-up assembly of smaller building blocks or the top-down assembly of a larger structure [56]. Bottom-up assembly exploits the structural features of the monomeric units that promote self-assembly and uses them to create larger structures [56, 57]. The process of self-assembly is found extensively throughout biological systems, for example cellular cytoskeletons, cilia and flagella in eukaryotic cells [57]. The cytoskeleton of cells is supported largely by actin filaments. The fibres measure up to microns in length and are formed from the specific head to tail association of actin [58]. In a similar system the association of tubulin forms microtubules, which make up the internal structure of cilla and flagella. Thus, nature provides us with a large database from which we can design new self-assembling biomaterials, either through copying and modifying existing systems or through generating new building blocks that have the potential to self-assemble [59].

As can be seen from the above discussion, peptides and proteins are lead candidates in this burgeoning field, whose structural characteristics can be exploited in the creation of biomaterials. The information needed to form the final 3D structure of a peptide or protein is contained within the primary sequence of amino acids. This means that interesting structural subunits can be chosen and used as the basis for a monomeric building block [59]. In addition, by studying the self-assembly of designed synthetic systems we may gain a greater understanding of how natural

systems self-assemble.

1.2.1 Self-Assembling Peptide Biomaterials

At present there are several different self-assembling systems that use peptide building blocks and have been extensively studied, these include: Amyloid-like structures [60]; α-Helical assemblies [61] and Peptide amphiphiles (PAs) [48, 62].

1.2.1.1 Amyloid-like Assemblies

The amyloid-like structures consist of peptides and proteins that have a propensity for forming β-strands; these strands can then assemble into sheets and eventually form fibrils [56]. Amyloid is commonly associated with diseases such as Alzheimer's [57] however, this scaffold has recently been engineered into a functional biomaterial by decorating the amyloid fibres with enzymes [63, 64]. In one example, Alkaline phosphatase and Horseradish peroxidase were attached via a linker to an amyloid-forming sequence. This oligomerised to produce a fibril with the enzyme decorating the surface of the fibril. These enzymes still displayed enzymatic activity when attached to the amyloid fibrils [63].

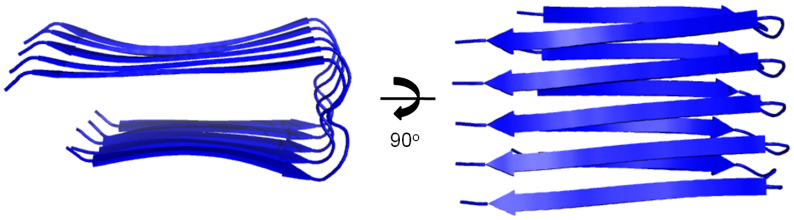

Figure 1.11: Structure of amyloid beta 1-42 fibrils (2BEG) [65].

1.2.1.2 Coiled coils

Coiled coil assemblies make use of α-helix-forming peptides that associate via favourable helix-helix interactions to form α-helical coiled coils. The motif consists of a heptad

repeat, where residues a and d in the sequence are non-polar, Figure 1.11. When 2 helices interact there is a hydrophobic interaction between the a and d residues, forming the main hydrophobic core. In addition, the residues adjacent to these (e and g) have an electrostatic interaction, which dictates the register between the helices. The propensity for the helices to self-associate has been exploited to create systems where components are linked together through a network of coiled coils, these include hydrogels [66] and fibres [67]. The hydrogel system used 2 complimentary heptad repeats attached to a hydrophilic polymer, when mixed in equal quantities the peptides formed a coiled coil crosslinking the polymer together to form a hydrogel [66]. In an interesting study, coiled-coils hydrogels have been shown to increase the proliferation and differentiation of neural stem cells and have the potential to be used for in vivo repair of nerve tissue [68]. Coiled coils fibres were produced by designing a pair of peptides that possess complementary charged groups in addition to the hydrophobic pair of residues. When mixed together, the peptides spontaneously self-assemble [67]. It was also possible to design spherical cages of coiled coils by combining different types of coiled coil peptide together [69].

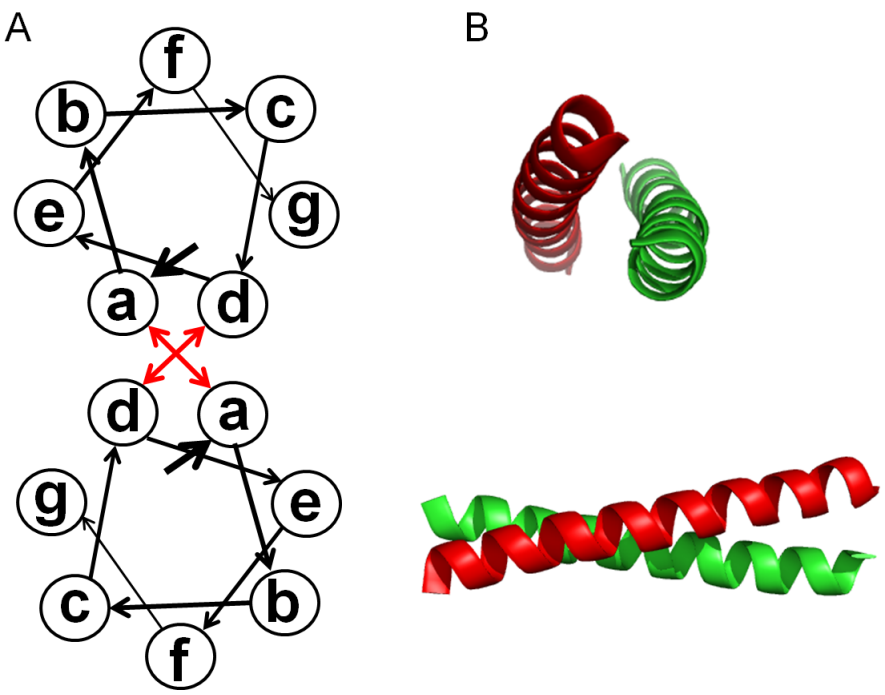

Figure 1.12: (A) Schematic of a dimer of 2 heptad repeats, showing how the 2 non-polar residues a and d form an interface between the 2 α-helices. (B) Structure of a coiled coil dimer (1ZIL).

1.2.1.3 Peptide Amphiphiles

In Peptide amphiphiles, a long alkyl chain is attached to the N-terminus of a peptide (Figure 1.13). These monomers assemble into micelles where the alkyl chains form a hydrophobic core and hydrophilic amino acids decorate the outside [56]. The shape of the self-assembled structure can be determined by altering the size and composition of the hydrophilic peptide group as well as through environmental factors [70]. Peptide amphiphiles have the potential to help form bioactive implants. For example, hybrid titanium-peptide amphiphiles implants were found to adhere more successfully to bone than the titanium implants alone [71]. The peptide amphiphiles present in the pores of the titanium allow growth of preosteoblastic cells inside the implant, leading to bone formation within the implant after 4 weeks [71]. Reduction in bone density at the interface between bone and implant is responsible for reducing the life-span of implants and therefore this technology has the potential

to improve the longevity of metal implants [71].

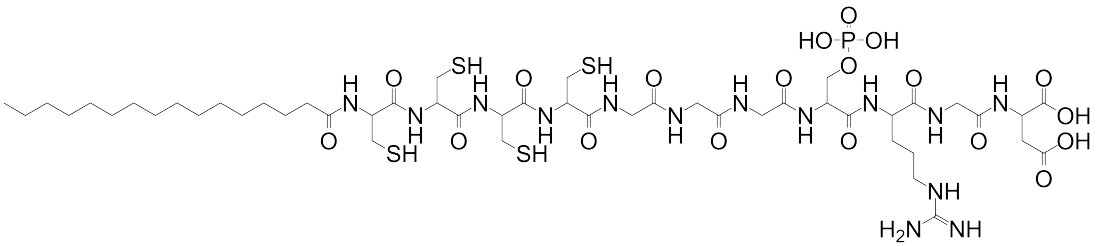

Figure 1.13: Structure of a peptide amphiphile monomer. There is an alkyl chain attached to a 11 residue peptide. This amphiphile reversibly forms a gel when treated with acid. [70]

1.2.2 Repeat Proteins as Building Blocks for Biomaterials

Repeat proteins provide a good structural basis from which to design novel self-assembling systems. The characteristic folds are often stable and rigid and are reproducible, even when the repeat units are isolated from larger structures [59]. Moreover, repeat proteins should provide excellent building blocks for self-assembly as they have the ability to be combined in a modular fashion to create novel and scalable chimeras/nanostructures with predictable structures, topologies, stabilities, linear length and function. For example, proteins produced from consensus TPR modules (CTPRs) can be recombinantly synthesised to form highly thermodynamically stable, very rigid rod-shaped superhelical arrays of up to 20 repeats [55, 72].

The driving force for self-assembly in the peptide based systems is the stabilisation energy that is gained on formation of an interface. The energetic gains from the formation of an interface between the repeat units is one of the driving forces of the folding of repeat proteins, where the intrinsic energies of the folding units are often less than the interface energy [73]. The importance of interactions between repeat interfaces was demonstrated by the spontaneously self-assembly of a pair engineered armadillo repeat fragments [74]. An armadillo repeat protein consisting of

3 internal repeats and an N and a C capping repeat was split into two components. When mixed together the N-terminal section spontaneously folds and associates with the C-terminal unit with a nanomolar K_d [74]. As it was possible to form dimers using the interfaces as the driving force, it is plausible that this mechanism could be exploited to drive the self-assembly of extended arrays of repeat modules.

Moreover, it has been shown that both the kinetic and equilibrium repeat protein folding is prone to the population of partially folded intermediate states [9,72,75,76]. Both the kinetic folding pathway and the population of the intermediate states can be changed by engineering changes in stability to individual repeats and/or the whole protein [75, 77–79]. This interesting feature could be used to create systems where the engineered unfolding of part of a repeat protein (for example a single terminal repeat) to leave a folded intermediate unit with a compatible oligomerisation interface. Thus such a system could respond to external stimuli, partially unfold and polymerise via compatible interfaces. However such a system has not yet been developed as most repeat proteins do not oligomerise even when compatible interfaces are present. One reason for this could be because the process to form ordered oligomers also competes with non-specific aggregation. At present other methods have been used to drive specific oligomerisation, such as metal chelation, peptide binding and covalent bonds.

1.2.2.1 Metal mediated Self-assembly

The β-roll motif often binds divalent metals (commonly Ca^{2+}) within the turn regions and, when tandemly arrayed, form a β-helical structure. Interestingly, if the divalent cations are removed they unfold and form disordered chains [44] [80]. This response to divalent metal ion concentration has been harnessed in two studies to produce novel triggerable fibre formation and a smart hydrogel [81, 82]. For fibre formation, Davies and co-workers designed minimized β-roll motifs that contained

either three β-strands and two Ca^{2+} binding sites (34 aa) or 5 β-strands and four Ca^{2+} binding sites (50 aa) [82]. Originally these constructs had been designed to fold into monomeric scaffolds on addition of Ca^{2+}. However, after being produced using peptide synthesis no structural change was observed on Ca^{2+} addition. However when lanthanum was added, filaments of approximately 3 nm in diameter and many nm in length were formed. The metal ion mediation oligomerisation was found to be reversible, as removing the La^{2+} through chelation with EDTA could disassemble the filaments. Some control of filament length was also obtained by introducing a further β-roll peptide that was capped at one end with biotin. In order to create a calcium responsive hydrogel Banta and co-workers also used a β-roll based protein [81]. Although here they created a chimera that possessed a leucine zipper domain N-terminally attached to modified β-roll array. The modified β-roll array was engineered to contain a leucine rich face. Once recombinantly produced, the leucine zipper/β-roll chimera form is soluble. However when calcium is added the β-roll arrays fold, exposing the leucine rich face. This causes the β-roll arrays to interact and oligomerise. The β-roll arrays oligomerisation, coupled with existing leucine zipper interactions, enable a weak hydrogel to form.

1.2.2.2 CTPR proteins as Biomaterials

Designed TPRs have been used to create both hydrogels [55], films [83] and fibres (as will be discussed in this thesis in Chapter 3). To form a hydrogel, an 18 repeat CTPR was used (termed CTPR18). This protein is extremely rigid and 18 nm long. Each CTPR18 was composed of alternating CTPR3 (a 3 repeat CTPR) units, one was a peptide binding CTPR3 and the other was a non-binding 'spacer' [55]. The CTPR18s were monomeric in solution, but could be triggered to create smart hydrogels or macroscopic solid films. To produce the hydrogels the CTPR18 proteins were incubated with a multivalent cognate peptide-PEG cross-linker. This caused multiple CTPR18s to be crosslinked together via the PEG-peptide component and

form a gel. The stimuli responsive 'smart' element was encoded through the peptide binding modules. When subjected to salt, the gel disassembles due to disruption of the electrostatic peptide/protein interaction.

Macroscopic solid films were formed by subjecting a solution of CTPR18 with 1 % w/v polyethylene glycol (PEG) 400 and leaving it to evaporate on Teflon tape [83]. Interestingly, this produced a 100 μM multilayered film that retains its α-helical content and functional ability to bind cognate peptide (when generated in its presence). X-ray scattering, coupled with birefringence of the film also suggested that the film possessed some macroscopic alignment and thus ordering of the CTPR18s.

1.2.3 Using covalent bonds to link monomeric units

In the examples we discussed in the previous sections, the monomeric units are not covalently linked together. The association is driven by a number of different mechanisms including: the formation of a hydrophobic core, the chelation to metals and protein-protein interaction. However, these systems often require that the conditions used for formation to be maintained in order to keep the assembly intact. For example the addition of salt to the CTPR18/peptide-peg crosslinker hydrogel becomes unstructured with the addition of NaCl, due to the disruption of the electrostatic effects. In order to form more robust structures that are tolerant to changes in conditions, the monomeric units could be covalently bonded together. There are a variety of methods for forming bonds between peptide/protein units. In Chapter 3 of this thesis we will discuss the use of a genetically encoded thioester that can undergo Native Chemical Ligation to link CTPR monomers covalently together, an example of Expressed protein ligation. However there are other methods available: for example isopeptide bonds and click chemistry.

1.2.3.1 Inteins for Expressed Protein Ligation or trans-splicing

Native Chemical ligation (NCL) is a reaction between two peptides that results in the formation of a peptide bond. This reaction requires an N-terminal cysteine and a C-terminal thioester. The thioester is required to increase the reactivity of the C-terminus, so that the cysteine can attack the carbonyl. Introducing a thioester to the C-terminus of proteins was difficult to achieve until the development of Expressed protein ligation (EPL) [84]. EPL exploits a naturally occurring protein splicing mechanism. The protein splicing mechanism involves the self-excision of an intein protein followed by the ligation between the C-terminal and N-terminal extein, Figure 1.14 Scheme A. In the first step of this process there is a slow N→Sacyl transfer, then nucleophilic attack occurs at the carbonyl. EPL removes the sequence that attacks the thioester, which leaves the protein open to attack from other nucleophiles, such as Mercaptoethanesulphonate (MESNa). The modified intein sequence can be combined with an affinity chromatography tag in order to combine purification and introduction of a thioester. Figure 1.14 Scheme B shows how the IMPACT purification system uses a modified intein sequence to produce proteins with a C-terminal thioester ready for ligation.

Intein trans-splicing uses an intein sequence that is split between the two fragments of protein to be joined together. When these fragments are mixed together the intein sequence catalyses the formation of a peptide bond between the C-terminal and the N-terminal fragment. This method has been used to ligate synthetic peptides with recombinant proteins. For example, GFP was labelled with the FLAG peptide through a trans-splicing reaction [85]. The GFP was expressed with the larger N-terminal section of the intein, the FLAG peptide was synthesised with the C-terminal remainder of the intein. When mixed together the GFP and the FLAG are ligated together and the intein sequence is excised.

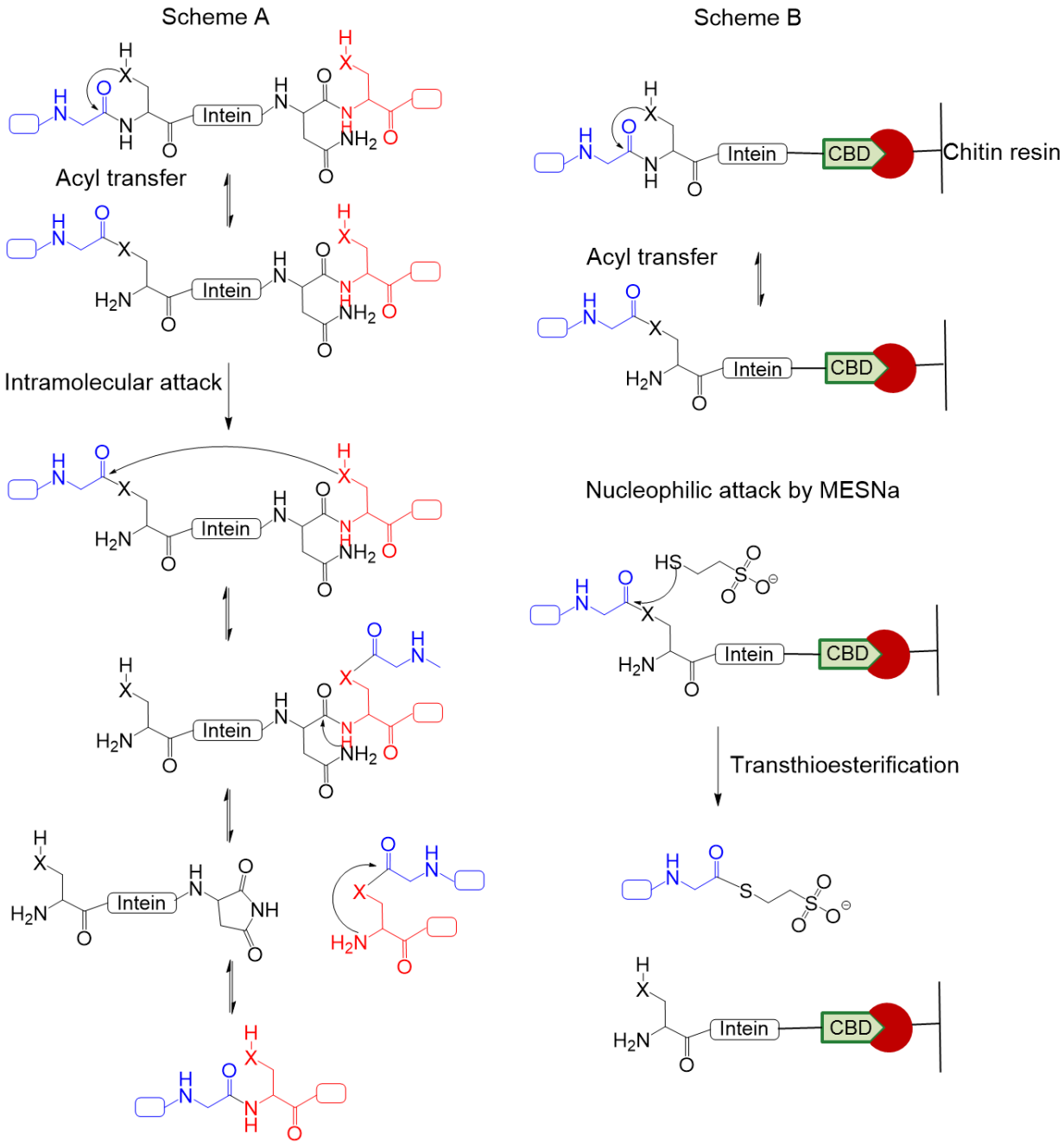

Figure 1.14: Scheme B: The full length fusion protein, target protein (blue) and intein and CBD (green), is bound via the CBD to chitin resin (red). The protein undergoes a spontaneous $S \rightarrow N$ acyl shift this generates a thioester. The thiol group of the reducing agent, Mercaptoethanesulphonate Sodium salt (MESNa), attacks the carbonyl group in a transthioesterification reaction. The protein is released from the resin and now possess a thioester at the C-terminus. The intein sequence remains bound to the chitin resin.

1.2.3.2 Using Isopeptide Bonds to Link Proteins

Isopeptide bond formation results in a covalent linkage between the amine group of a lysine residue and the carboxylate group of an aspartate or glutamate or the amide group of an asparagine. One example of this type of linkage is found in the surface phili of the Gram positive bacteria *Streptococcus pyogenes* [86]. Here the isopeptide formation between an aspartate and a lysine residue is catalysed by a nearby glutamate residue within a domain comprised of β sheets.

The discovery of this naturally occurring example has led to the re-design and optimisation of this system to form the SpyCatcher/SpyTag pair [87], Figure 1.15. This redesign involved splitting the protein into two components: the bulk of the protein was the SpyCatcher (138 residues) and the smaller fragment became the SpyTag (13 residues). This protein-peptide pair spontaneously form an isopeptide bond between a lysine on the SpyCatcher and an aspartate on the SpyTag. This reaction is rapid and results in the formation of the isopeptide bond between 40 % of the protein/peptide pairs within the first minute of reaction. Once reacted the pair is very stable, and remains folded down to pH 2 and up to a temperature of 100 °C. The tolerance to such a diverse range of conditions presents the possibility of utilising this reaction as a labelling or protein purification system.

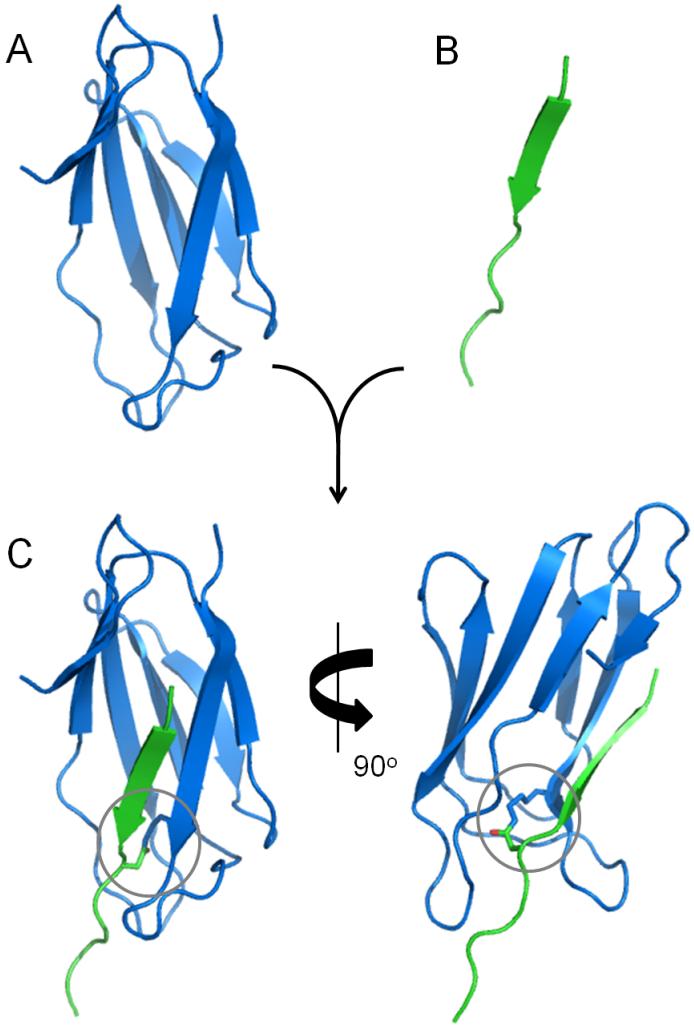

Figure 1.15: Crystal structure of SpyCatcher (A) and SpyTag (B) (4MLI). When mixed together they form complex (C), that is held together by hydrogen bonding between the β sheets and an isopeptide bond (circled) between a lysine residue on the SpyCatcher and an aspartate residues on the SpyTag.

This system was developed further to create SpyLigase [88]. Here the requirement for a catalytic glutamate residue for isopeptide bond formation was exploited and the SpyCatcher protein was divided into: SpyLigase (containing the glutamate residue) and KTag (containing the lysine residue). When all 3 components were mixed together the SpyLigase mediated the isopeptide bond formation between the SpyTag and the KTag peptides.

In the groups most recent work, SpyTag and SpyCatcher were combined with dead

streptavidin (a non-biotin binding variant of streptavidin) to create SpyAvidin [89]. These were then mixed with Traptavidin (a stronger binding variant of streptavidin) to form tetramers that contained mixtures of the Traptavidin and the SpyAvidin. Complex structures were formed by isolating selected isomers and then reacting the SpyTag variant with the SpyCatcher variant. An octamer, capable of binding up to 6 biotin ligands was formed from two tetramers: one containing 3 traptavidin units and a dead streptavidin with a SpyTag and one containing 3 traptavidin units and a dead streptavidin and a SpyCatcher. When these 2 tetramers are mixed together they form an isopeptide bond between the SpyTag/SpyCatcher pair. By mixing different tetramers together it was possible to form an Eicosamer (20 subunits). The Eicosamer is capable of binding 12 biotinylated MHC units, which interact with T-cell receptors on T-cells [89].

In another study fibrils were formed by re-engineering the SpyCatcher/SpyTag system to promote directional polymerisation [90]. The Spytag peptide was transferred from the C-terminus to the N-terminus of the Spy protein, leaving a binding pocket that can accommodate the spytag from another monomer. The binding pocket was protected from reaction by a capping peptide tag, held in place by a disulfide bond. When the disulfide was reduced the pocket was free to bind to the Spytag and was bonded irreversibly by the formation of an isopeptide bond [90].

1.2.3.3 Using Click chemistry to link proteins

Click chemistry is a reaction that produces a covalent carbon-heteroatom bond. Although this is not one specific reaction it is used to describe a reaction that: produces high yields, can be performed under mild conditions (i.e. in water) and where any by-products can be removed without using chromatography [91]. The most widely used 'Click' reaction is a Huisgen [3+2] cycloaddition between an azide and an alkyne, see Figure 1.16.

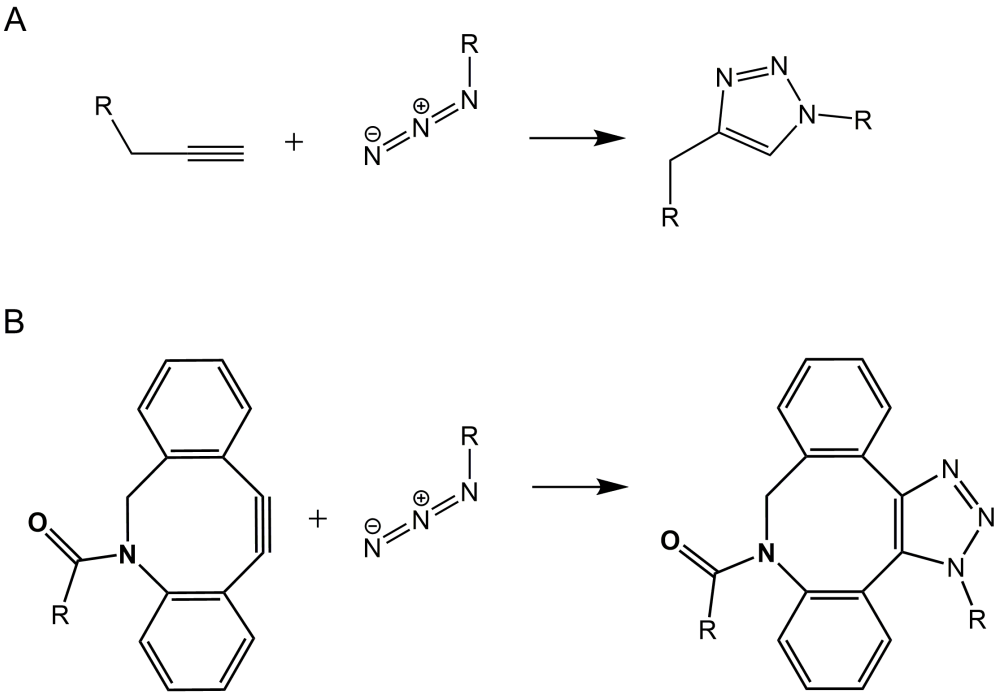

Figure 1.16: An example of a [3+2] azide-alkyne cycloaddition

The example shown in Figure 1.16 A requires a Cu(I) catalyst in order to react at low temperatures. Copper-free cycloadditions have been developed by increasing the reactivity of the alkyne. By placing the alkyne as part of a strained ring system the reactivity is increased and the reaction can proceed without a catalyst. As click chemistry reactions can be performed under mild conditions in water they have been used for a variety of bioconjugation reactions involving proteins. As well as the favourable reaction conditions the reactive groups required for click chemistry are orthogonal to the functional groups possessed by canonical amino acids [92].

Click chemistry has been used to attach spectroscopic tags to proteins. For example, the Cowpea mosaic virus (CPMV) was decorated with fluorescein. The virus protein was functionalised with azides or alkynes via reaction with surface lysine or cysteine residues. Fluorescein was linked to the protein via the complementary azide or alkyne [93]. Achieving the fluorescent labelling via several routes demonstrates the versatility of the click cycloaddition in protein decoration.

Click chemistry-compatible non-natural amino acids can now be genetically programmed into proteins. Site-specific incorporation is achieved by using an orthogonal pair of t-RNA synthetases and the amber codon, TAG [94]. To create proteins using this method, they are grown in media that includes the non-natural amino acid and when the amber stop codon is encountered, the non-natural amino acid is inserted by the orthogonal t-RNA. Pluncktun and co-workers developed a tumour-binding DARPin into one that was labelled using Click Chemistry [95]. The DARPin was expressed using a system that incorporated the non-natural amino acid azidohomoalanine at the N-terminus, in place of methionine. This azide reacts with a cyclooctyne without the need for a copper catalyst, which would be toxic to cells. This reaction is specific and orthogonal to any other side reactions, so only the target azide reacts with the clickable reagent. The result is that the DARPin can be specifically tagged at the N-terminus with polyethylene glycol (PEG). The advantage of functionalising proteins with PEG is that it has been shown to increase the serum persistence of proteins, however this addition can decrease the binding ability of the protein. This DARPin was also functionalised at the C-terminus with a fluorophore tag, which was linked through a maleimide linkage to a cysteine. The addition of the fluorophore allowed more detailed studies of the binding kinetics, where PEGylation was found to decrease binding rate but not the dissociation rate.

1.3 Aims

The aim of this thesis is to develop the CTPR scaffold into a self-assembling biomaterial. In Chapter 3 we engineered a CTPR protein that, when triggered, would polymerise. Repeated reaction between the monomers/oligomers lead to the formation of fibres. We began the development of this system by investigating the properties of a 3 repeat CTPR that lacks the C-terminal helix. This C-terminal helix was originally engineered into the CTPR proteins as a capping helix, to reduce aggregation. This CTPR3ΔS was then developed into a system that could be triggered to undergo Native Chemical Ligation, using a genetically encoded intein domain that is cleaved using the thioester, Mercaptoethanesulfonate.

In Chapter 4 we investigated the stability of deletion mutants of 2 series of CTPR proteins. These mutants were designed to lack either the N-terminal or C-terminal helix, i.e. half a repeat unit. This data was used to develop a Heteropolymer Ising model, where the contribution to the stability of the protein was assessed for the terminal helices. The knowledge gained was used to model a chimera of the 2 protein series containing the stabilising inter-repeat loop sequence `PNN` in the main body of the protein and the less stable sequence `PRS` in the last loop before the C-terminal helix. The aim of this was to create a protein that could, under certain conditions, populate a partially folded intermediate. The goal of this is to develop a system that can partially unfold and then oligomerise.

In Chapter 5 we explore using a photoisomerisable linker to disrupt the stability of the C-terminal helix. This linker is attached to the helix via 2 cysteine residues and when exposed to UV light, isomerises from trans to cis.

Chapter 2

Methods and Materials

2.1 Vectors

The genes encoding for the various CTPR proteins used in this thesis were inserted into the multi-cloning site of either pProEx HTb, pBAD HisA or pTrcHis-TOPO (Figure 2.1) The pProEx HTb vector was used to express and purify CTPR proteins used in Chapters 3, 4 and 5. When used this produced a CTPR protein with an N-terminal His-Tag. The pTrcHis-TOPO was used to express CTPR proteins for intein mediated protein polymerisation used in Chapter 3. At the start of this thesis Dr E. Main, Dr J.J. Philips and Dr Y. Javadi provided a number of CTPR genes in either the pProEX HTb or pTrcHis-TOPO vector (Table 2.1).

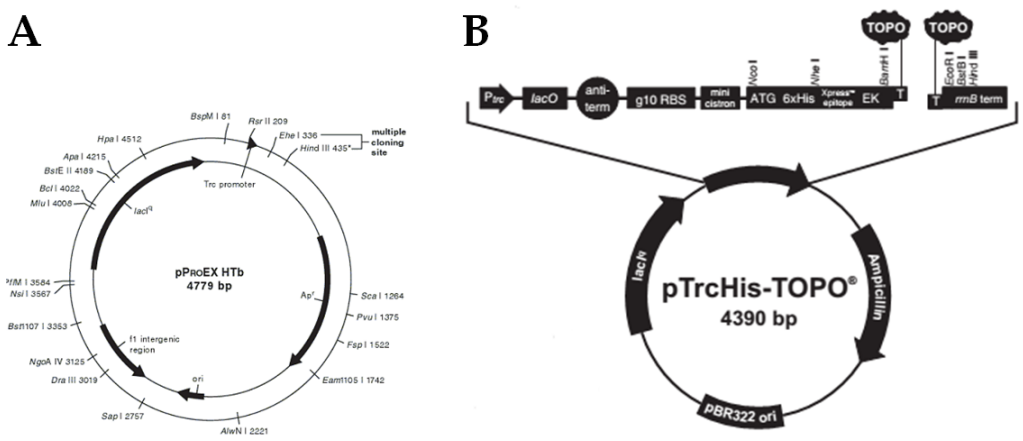

Figure 2.1: Vectors used for expression of recombinant proteins during this thesis. (A) pProEX HTb. This was used for expression of CTPR proteins, the gene was ligated between the BamHI and HindIII sites. Features include an ampicillin resistance gene, allowing for selection of cells that contain the plasmid and the Trc promoter and LacI sequence allows for over-expression of the protein using isopropyl β-D-1-thiogalactopyranoside (IPTG). Vector map from Invitrogen Life Technologies manual [96]. (B) pTrcHis-TOPO. This vector was used in the expression of the CTPR3ΔS-Intein-Chitin Binding Domain. The vector contains the ampicillin resistance gene. The LacI sequence allows the protein to be over-expressed using IPTG. Vector map of B from Invitrogen Life Technologies instruction manual [97]

Table 2.1: Plasmids available at the start of the project

Vector	**Gene**
pProEX HTb	CTPR2
	CTPR3
	CTPR3ΔS[2]
	CTPRa2[1]
	CTPRa2ΔS[12]
	CTPRa3[1]
pTrcHis TOPO	CTPR3ΔS-Intein-Chitin Binding Domain[2],[3]
[1]CTPRan proteins (n=number of repeats) the loop sequence between the repeats has been changed from PNN to PRS	
[2]ΔS mutants do not possess the C-terminal helix. This was achieved by placing a stop codon at position 73 in CTPR2 and 106 in CTPR3.	
[3]The intein domain allows for cleavage of the protein to be induced using a reducing agent. The chitin binding domain allows for affinity purification using chitin resin.	

2.1.1 Mutant CTPR Proteins

In order to undertake the work in this thesis a number of CTPR proteins were cloned using the genes listed in Table 2.1. These new CTPR proteins are summarised and described in Table 2.2.

Table 2.2: Mutant CTPRs constructed during this Thesis

Gene based on	Name of Mutant	Vector Used	Methods Section	Construction Method
CTPR2	CTPRy2[1]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR2ΔS[2]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR2ΔA[3]	pProEX HTb	Section 2.2.1	PCR
CTPRa2	CTPRa2ΔA[3]	pProEX HTb	Section 2.2.1	PCR
CTPR3	CTPRy3[1]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPRa3ΔA[3]	pProEX HTb	Section 2.2.1	PCR
	CTPR3i4[4]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR3i11[5]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR3yi4[4]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR3yi11[5]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPR3yi11 I97C[5]	pProEX HTb	Section 2.2.3	Round-the-horn PCR
CTPRa3	CTPRa3ΔS[2]	pProEX HTb	Section 2.2.2	Site-directed Mutagenesis
	CTPRa3ΔA[3]	pProEX HTb	Section 2.2.1	PCR

[1] The final loop before the C-terminal helix has been mutated from PNN to PRS, at positions 70-72 in CTPR2 and 103-105 in CTPR3.

[2] ΔS mutants do not possess the C-terminal helix. This was achieved by placing a stop codon at position 73 in CTPR2 and 106 in CTPR3.

[3] ΔA mutants do not possess the N-terminal helix and the 3 residue loop before the start of the second helix. The first 16 residues have been removed.

[4] i4 mutants contain cysteine mutations at positions 115 and 125.

[5] i11 mutants contain cysteine mutations at positions 111 and 122.

2.2 Construction of CTPR mutants

2.2.1 Construction of CTPR mutants without the N-terminal helix

2.2.1.1 Polymerase Chain Reaction (PCR) Amplification of CTPR gene

Polymerase Chain Reaction (PCR) amplification was used to create truncated versions of the CTPRn and CTPRan proteins, where the N-terminal α-helix is absent. The gene that was created with this method contained a BamHI restriction site at the 5' end and a HindIII restriction site at the 3' end. This was to allow ligation between the corresponding sites in the pProEX HTb vector. Four mutants were created in total: CTPR2ΔA, CTPRa2ΔA, CTPR3ΔA, CTPRa3ΔA. A schematic of the primer design for this type of PCR is shown in Figure 2.2. Primers were designed to have a melting temperature of between 60-65 °C and be the same length (± 1-3 bases). The primers were 33 and 34 bases in length. Primers were purchased from Integrated DNA Technologies. The forward primer anneals where the sequence encodes for the residues at position 16, in order to remove the N-terminal helix and the 3 residue loop before the start of the second helix. A BamHI restriction site was incorporated into the coding primer at the 5' end, which does not anneal to the template sequence (shown in red in Figure 2.2).

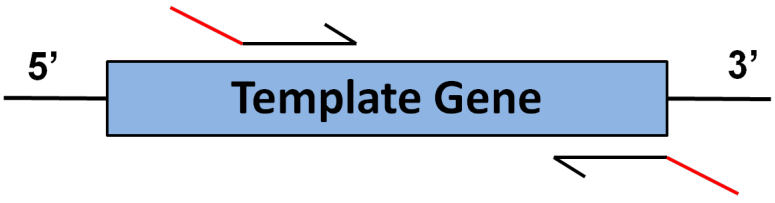

Figure 2.2: Schematic showing the where primers were designed to anneal to the template sequence. Restriction enzyme sites were incorporated at the ends of the primers (shown in red). A BamHI site was added to the forward primer and HindIII was added to the reverse primer.

The reverse primer anneals where the sequence encodes for the C-terminus of the CTPR. A HindIII restriction site was added to the 5' end of the primer. The restriction enzyme sites were incorporated into the ends of the primers to allow digestion and ligation into the pPROEX HTb vector.

The PCR reaction mixture contained the following: 5 μl 10X Reaction buffer, 2 μl template DNA, 1 μl Forward primer (50 μM stock), 1 μl Reverse primer (50 μM stock), 1 μl dNTPs (10mM stock), 39 μl ddH_2O and 1μl Polymerase (typically Taq or Pfu Turbo) to give a total volume of 50 μl. The reaction was cycled through the temperatures shown in Table 2.3.

Table 2.3: PCR program for reaction described in Section 2.2.1

Segment	Temperature	Time
1	95°C	60 s
2*	95°C	30 s
	50-55°C	30 s
	68-72°C	60 s
3	68-72°C	10 minutes

*Segment 2 was repeated 25 times.

2.2.1.2 Agarose Gel Electrophoresis

Following the PCR, the success of the reaction was judged using electrophoresis on a 1 % (w/v) Agarose gel (preparation is described in the Appendix Section 7.0.4. The running buffer was 1x Tris-Borate EDTA (TBE). DNA samples were mixed with 6X Loading Buffer (30% Glycerol, 0.25% Bromophenol Blue) before loading onto the gel. The running voltage was between 60-100 V, and the voltage was applied until the dye had progressed 2/3 of the way across the gel. DNA bands were visualised using UV at 230 nm.

2.2.1.3 Restriction Enzyme Digestion

After a successful PCR reaction, the gene and pProEX HTb vector was digested with BamHI and HindIII to produce the correct complementary overhanging ends to allow ligation.

The digest mixture was as follows: 16 μl DNA (PCR reaction mixture, vector), 2 μl 10X reaction buffer, 1 μl BamHI and 1 μl HindIII to a volume of 20 μl. The reaction mixture was incubated for 1 hour (for digestions performed with NEB enzymes) or between 5-60 minutes (when using Fermentas Fastdigest enzymes).

2.2.1.4 Dephosphorylation of Digested pProEX HTb

To prevent the possibility of false positives caused by vector re-ligation, digested pProEX HTb were dephosphorylated prior to ligation with the inserts. The following was a typical reaction mixture: 16 μl Digested Vector, 2 μl 10X reaction buffer and 1 μl Alkaline Phosphatase to a volume of 20 μl. The reaction was incubated at 37 °C for 10 minutes, followed by deactivation at 75 °C for 5 minutes.

2.2.1.5 Gel Purification of Digested Vectors and Inserts

The digested pPROEX HTb and inserts were purified using agarose gel electrophoresis. The agarose gel was prepared as described in Appendix 1 Section 7.0.4. The concentration of the Agarose was varied according to the length of the DNA, with smaller fragments <500bp requiring a higher concentration (1.5-2 % w/v) and larger fragments >4kb requiring a lower concentration (0.7 %)

DNA bands were visualised using UV at 230 nm. Where a band of the correct size was observed, it was cut from the gel using a scalpel. The DNA was recovered from the gel using a Fermentas GeneJet Gel Extraction kit. The suggested protocol

was followed [98], except the final elution was achieved using 20 μl ddH_2O. DNA was stored at -20 ^oC after purification.

2.2.1.6 Ligation of Purified Vector and Inserts

Ligation reactions were performed using the following volumes: 8 μl DNA (1 μl vector and 7 μl insert), 1 μl 10X Reaction buffer and 1 μl T4 DNA Ligase to a give a total volume of 10 μl. A ratio of 1:7 (vector: insert) was used. The reaction was incubated at 4 ^oC overnight or 16 ^oC for 2 hours.

2.2.1.7 Transformation of Ligation reaction into E.coli Cells

Ligations were transformed into electrocompetent XL2 Blue Ultracompetent E.coli cells from Agilent Technologies. Electrocompetent cells were kept frozen at -80 ^oC in aliquots of 50-100 μl until required. To perform a transformation an aliquot of cells was thawed on ice and 1 μl of ligation reaction mixture was added and placed in a 0.2 mm MicroPulsar cuvette. A Biorad Genepulsar, set to bacteria, was used to electroshock the cells. A successful pulse gave a reading of $>$ 4ms. Cells were recovered with 1 ml of LB broth and incubated for 45 minutes at 37 ^oC. After incubation, 20-100 μl of the cells was spread onto L-Agar containing Ampicillin (0.1 mg/ml). The plate was incubated overnight at 37 ^oC. If the transformation was successful, 10-200 discrete colonies will be seen on the plate.

2.2.1.8 Extraction of DNA from E.coli

Single colonies were picked and grown overnight in LB Broth. The cells were harvested and a DNA extraction was performed using either a Qiagen QiaJet or Fermentas GeneJet plasmid purification kit. The suggested protocol in the Qiagen QiaJet [99] or Fermentas GeneJet [100] manual was followed except the DNA was eluted with 100 μl ddH_2O.

2.2.1.9 DNA Sequencing

To confirm the success of a cloning reaction, a sample of the plasmid (15 μl at a concentration of 100 ng/μl) was sent for sequencing to Beckman Coulter Genomics.

2.2.2 Creating 1-4 Base Mutations in CTPRn and CTPRan using Site-directed Mutagenesis

Site-directed mutagenesis was used to generate small mutations of between 1 and 4 bases. This method was used to create the CTPRyn, CTPRnΔS and the double Cysteine mutants CTPR3i4, CTPR3i11 and CTPRy3i11 (Table 2.2). Primers were designed as directed by the Stratagene Quik-Change kit [101] (see Figure 2.3) and were purchased from Integrated DNA Technologies (IDT). For each mutation the primers were designed to be between 25-45 bases in length. In this technique the primers are designed so that the mutation is included as a loop between 2 annealing sections, the mutation is shown in red in Figure 2.3, and the annealing portions are shown in black.

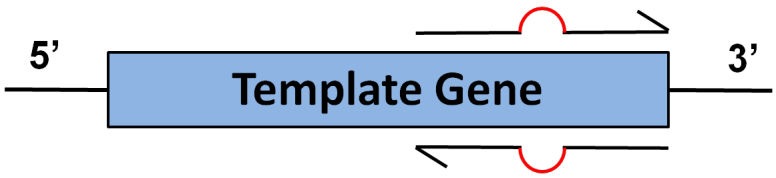

Figure 2.3: Schematic showing where the primers for Site-directed mutagenesis anneal to the template gene. The primers are designed with the mutation site in the centre of the primer (shown in red)

A typical reaction mixture contained: 5 μl 10X Pfu buffer, 2 μl template DNA, 1.25 μl Forward primer (10μM), 1.25 μl Reverse primer (10μM), 1 μl dNTPs (10mM) 38.5 μl ddH_2O and 1 μl Pfu Turbo DNA Polymerase to give a total volume of 50 μl.

The reaction was cycled through the temperatures shown in Table 2.4 The annealing temperature was varied between 50-55 °C according to the melting temperature of the primers.

Table 2.4: Site-directed mutagenesis PCR temperature cycle

Segment	Temperature	Time
1	95°C	60 s
2*	95°C	50 s
	50-55°C	60 s
	68°C	7 minutes
3	68°C	10 minutes
*Segment 2 was repeated 18 times		

Following IPCR, DpnI was used to digest the methylated parental strands of DNA (from the plasmid template). 1 μl of DpnI was added to the 50 μl reaction mixture and incubated at 37°C for 1 hour. 1 μl of this mixture was transformed into XL2 Blue E.coli cells as described in Section 2.2.1.7. The DNA was extracted and sequenced as described in Sections 2.2.1.8 and 2.2.1.9.

2.2.3 Round-the-horn Mutagenesis

As an alternative to Site-directed mutagenesis, Round-the-horn mutagenesis was used to create small mutations of 1-5 bases. Like site-directed mutagenesis, this technique is also a type of Inverse Polymerase Chain Reaction (IPCR), where the whole plasmid is amplified.

This technique was used to create CTPRy3i11 I97C (Table 2.2). Primers were designed to start from the desired mutation site, with the mutated sequence included at the 5' end of the forward primer, shown in red in Figure 2.4.

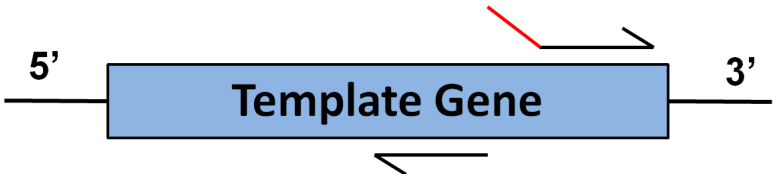

Figure 2.4: Schematic showing where the primers for Round-the-horn mutagenesis anneal to the template gene. Primers are designed so the mutation is at the 5’ end of the Forward primer (shown in red). After the PCR, blunt ended linear strands of DNA containing the mutation are produced.

2.2.3.1 Phosphorylation of Primers for use in Round-the-horn mutagenesis

Round-the-horn mutagenesis PCR produces linear, blunt-ended DNA that needs to be ligated before transformation. To ligate linear DNA the ends need to contain phosphate groups. The primers were purchased without phosphorylated ends, therefore these were added using the following protocol. A reaction mixture containing the following was prepared: 5 μl Primer (100 μM Stock), 5 μl 10X Reaction buffer, 1 μl 50mM $MgSO_4$, 37 μl ddH_2O, 1 μl ATP (100mM stock) and 1 μl PNK to give a total volume of 50 μl. The reaction was incubated at 37 ^{o}C for 1 hour followed by deactivation at 65 ^{o}C for 10 minutes.

2.2.3.2 IPCR of Phosphorylated Primers

After phosphorylation of the primers, the mutant gene was amplified using the following reaction mixture: 5 μl 10X reaction buffer, 2 μl template DNA, 1.5 μl Phosphorylated Forward primer (10 μM), 1.5 μl Phosphorylated Reverse primer (10 μM), 1 μl dNTP mix (10mM stock), 38 μl ddH_2O and 1 μl Pfu Turbo DNA Polymerase to give a total volume of 50 μl. The reaction was cycled through the temperatures shown in Table 2.5. After the IPCR, the DNA was gel purified and

Table 2.5: Round-the-horn mutagenesis PCR temperature cycle

Segment	**Temperature**	**Time**
1	95 °C	60 s
2*	95 °C	30 s
	50-58 °C	30 s
	72 °C	12 minutes
3	95 °C	60 s

*Segment 2 was repeated 25 times

ligated as described in Section 2.2.1.5. The DNA was transformed, extracted and sequenced as described in Sections 2.2.1.7 to 2.2.1.9.

2.3 Protein Expression and Purification

2.3.1 Media for Culturing E.coli

Two types of media were used: LB Broth and 2XYT. LB broth was purchased from Melford and dissolved according to the instructions (20 g/L). 2XYT was also purchased from Melford and dissolved according to the instructions (32 g/L). Media was always dissolved in dH_2O and autoclaved the same day. LB broth was used for overnight starter cultures and 2XYT was used for protein expression.

2.3.2 Expression and purification of CTPRn and CTPRan proteins from pProEX HTb

2.3.2.1 Expression of recombinant CTPR protein

The plasmid was transformed into the E.coli expression cell line OverExpress C41 (DE3) (Lucigen) as described in Section 2.2.1.7. 10ml starter cultures were grown from single colonies overnight at 37 °C. Each 10 ml of starter culture was added to a 2 L flask (containing 1 L of 2YT media and 0.1 mg/ml Ampicillin). The culture was incubated at 37 °C with shaking at 210 rpm until the optical density at 600 nm was between 0.6-1.0. A final concentration of 100 μM IPTG was added to induce expression and the cells were grown for a further 3-5 hours at 30 °C, or 18-25 °C overnight. The cells were harvested by centrifugation (5 minutes at 7000 rpm) and re-suspended in phosphate buffered saline (1xPBS) pH 7. The cells were either flash frozen (for purification at a later date) or lysed immediately.

2.3.2.2 Cell Lysis

Cells were lysed using either an Emulsiflex or Sonication (5 minutes at 30 s on 30 s off). Cell debris was removed from the soluble protein through centrifugation at 17 000 rpm for 1 hour. The supernatant contained the protein of interest.

2.3.3 Affinity Purification of His-tagged fusion CTPR proteins

All pProEX HTb proteins were expressed as fusions with an N-terminal Histidine tag. To purify the protein, a Nickel-containing agarose resin was used. Approximately 10 ml of resin was used for every 4 L of culture grown. The Ni-containing agarose was produced by charging iminodiacetic acid resin (Sigma-Aldrich) with 100 mM $NiSO_4$ and washed with 2-3 column volumes (CV) of 1xPBS pH 7. The crude mixture of soluble protein was either bound by gravity flow or batch bound by shaking the resin and protein for 1 hour. The protein bound resin was washed with at least 250 ml of 1xPBS. Protein was eluted using 1-5 CVs of 250 mM Imidazole in PBS at pH 7. To cleave the His-tag, TEV protease was added to the eluted protein and incubated overnight at room temperature. The resin was cleaned after use by washing with 2-3 CV of 100 mM EDTA, 4-5 CV ddH_2O and finally 2-3 CV of 20 % Ethanol before being stored at 4 °C.

2.3.4 Gel Filtration of Cleaved CTPR Proteins

A HiLoad 26/600 Superdex 75 pg column (GE Healthcare) was used to purify the proteins after affinity chromatography and cleavage of the His-tag. The column was attached to an Akta FPLC machine with an automated fraction collector. A flow rate of 3 ml/minute was used. The column was equilibrated with 2 CV of PBS pH 7 and the concentrated protein (volume 10-20 ml) was loaded via a superloop. Elution

of the protein was monitored using UV absorbance at 280 nm. The fractions were analysed using SDS-PAGE, and fractions containing the desired protein were pooled and concentrated.

2.3.5 Buffer Exchange

Proteins were buffer exchanged using dialysis or desalted with a PD-10 column as follows.

2.3.5.1 Dialysis

Snakeskin dialysis membrane was purchased from Fisher. The Molecular Weight Cut Off (MWCO) used was 3000 Da. The tubing was cut to the desired size and wet with the new buffer. A 4 L volume of the new buffer was prepared. The protein was loaded into the dialysis membrane, with clips at each to prevent leakage. The dialysis was performed overnight with stirring at 4 °C.

2.3.5.2 Desalting Columns

A PD-10 desalting column was used to buffer exchange protein volumes of less than 2.5 ml. The column was equilibrated using gravity flow with 25 ml of the new buffer. The protein was applied to the column in a 2.5 ml aliquot. Once the protein had fully entered the column, 3.5 ml of the elution buffer was added to elute the protein.

2.3.6 Storage of CTPR Proteins

Purified proteins were concentrated to a minimum of 100 μM and flash frozen and stored at -80 °C. The buffer was either 1xPBS or 50 mM phosphate at pH 7.

2.3.7 N^{15} Labelled CTPR Proteins from pProEX HTb

N^{15} labelled proteins were prepared from the pPROEX HTb vector to allow 2D NMR data to be collected.

2.3.7.1 N^{15} Enriched Media

To produce N^{15} protein, the E.coli was cultured in media where the only nitrogen source is enriched N^{15}. The following media recipe was used. A litre of M9 minimal media was prepared by autoclaving 950ml of ddH_2O containing: 6 g Na_2HPO_4, 3 g $KHPO_4$, 5 g NaCl and 10ml of 100X trace element solution. A 50 ml in ddH_2O additive was prepared containing:1 mg thiamine, 1 mg biotin, 20 mM $MgSO_4$, 8 mM $CaCl_2$, 1 g $^{15}NH_4Cl_2$ and 4 g Glucose.

2.3.7.2 Expression of N^{15} Labelled Proteins

The expression and purification of the N^{15} labelled proteins follows the same procedures as those described in Section 2.3.3 with the following exceptions described below:

A starter culture of 25 ml was pelleted and re-suspended in 5 ml 1xPBS pH 7 and added to 1 L of M9 minimal media, enriched with N^{15}. This was grown at 37 ^{o}C for 6-10 hours until an OD of between 0.6 and 1.0 was reached. 100 μM IPTG was added and the culture incubated at 18 ^{o}C overnight.

2.3.8 Expresssion and Purification of the His-Xa-CTPR3ΔS-Intein-Chitin binding domain protein

The pTrcHis-TOPO vector protein was expressed as a fusion with an N-terminal His-tag and a C-terminal chitin binding domain (His-Xa-CTPR3ΔS-Intein-Chitin binding domain in pTrcHis-TOPO vector). The same expression protocol was followed as described in Section 2.3.2. The cells were lysed as described in Section 2.3.2.2.

Chitin resin (New England Biolabs) was used to purify the protein. The soluble protein fraction was bound by gravity flow. The resin was washed with 25 CVs of 1xPBS pH 7. Elution was accomplished by adding 1.5 CVs of cleavage solution (0.5 M MercaptoethansulphonateNa (MESNa), 50 mM Tris, 150 mM NaCl, and 1 mM EDTA at pH 7) and incubating overnight with rotating at room temperature. The cleaved protein was washed off the column with a further 2-3 CV of cleavage solution. The reducing agent MESNa cleaves the protein from the resin and leaves the chitin binding domain attached to the column and generates a thioester at the C-terminus of the fusion protein to produce the construct: His-Xa-CTPR3ΔS-MESNa. The protein was further purified using Size Exclusion Chromatography as described in Section 2.3.4 and the purified protein analysed using SDS-PAGE.

The Chitin resin was regenerated by washing with 4-5 CV of 0.3 M NaOH, 4-5 CV of ddH_2O and was stored in 20 % Ethanol at 4 ^oC.

2.4 Protein Purity and Identity

2.4.1 Protein Concentration Determination

The concentration of purified proteins was calculated using the Beer-Lambert Law. A UV spectrum was measured between 200-400 nm and the absorbance at 280 nm was used with equation 2.1 to calculate the concentration.

$$A = \varepsilon cl \tag{2.1}$$

Where A is the absorbance at 280 nm, ε is the Molar Extinction coefficient($M^{-1}cm^{-1}$), c is the concentration (in M) and l is the pathlength of the cuvette. The Molar Extinction Coefficient was calculated for each protein using Equation 2.2 [102]. This is based on the number of tryptophan, tyrosine and cysteine residues and the extinction coefficient for each of these residues.

$$\varepsilon = (\text{no. Tyr})(1490) + (\text{no. Trp})(5500) + (\text{no. Cys})(125) \tag{2.2}$$

2.4.2 Polyacrylamide Gel Electrophoresis (PAGE)

2.4.2.1 SDS-PAGE Analysis of Proteins

Denaturing SDS-PAGE gels were used to assess the purity of protein samples. SDS-PAGE gels used during this thesis were 18% and were prepared as described in Appendix 1, Section 7.0.4.1. Gels were run at a constant voltage of between 150-200 V until the dye front had reached the bottom of the gel.

To visualise the bands, gels were stained with Coomassie brilliant blue (2.5 g Coomassie, 450 ml Methanol, 450 ml ddH_2O and 100 ml Acetic acid). The gel was submerged

and heated for 1 minute and incubated with shaking for 10-30 minutes. Gels were de-stained using the following solution: 10 % (v/v) Acetic acid and 10 % (v/v) Methanol in ddH_2O. The gel was incubated with shaking until sufficiently de-stained.

2.4.2.2 Native PAGE

Native PAGE gels were prepared in the same manner as non-native except SDS was substituted with water in the buffer recipes. DTT was removed from the loading buffer.

2.4.3 Mass Spectrometry

Purified proteins were analysed using MALDI-TOF mass spectrometry on a Bruker Daltonics Autoflex. Samples were first de-salted using a Millipore C18 ZipTip before being mixed with a matrix of sinapinic acid. The matrix was made by saturating acetonitrile with sinapinic acid. A 3 μl sample of matrix and protein samples (1:1 ratio) was spotted onto the MALDI sample plate and allowed to air dry before being loaded into the machine. The Laser power and gain was varied until the best signal output was observed.

2.5 Equilibrium Unfolding Studies of CTPR Proteins

Analysing the reversible equilibrium unfolding of a protein yields information about the thermodynamic stability. Here, the unfolding equilibria of the CTPR proteins were induced by either changing the temperature (thermal) or by addition of the chemical denaturant GuHCl. The unfolding reactions were monitored through the change in the proteins in Fluorescence or ellipticity. Data collected was analysed using the equations described in Section 2.9.

2.5.1 Probes to Measure CTPR Unfolding

2.5.1.1 Circular Dichroism

Far UV Circular Dichroism monitors the secondary structure of proteins. CTPR proteins are very α-helical and this structural feature yields a negative peak for the ellipticity at 222 nm. When the CTPR is unfolded it loses its α-helical structure and the ellipticity increases until it reaches zero, at this point the protein is fully unfolded.

2.5.1.2 Fluorescence

Fluorescence can be used to monitor the tertiary structure of a protein that contains tryptophans and tyrosines. The protein samples is excited at 280 nm (the absorbance of tryptophan and tyrosine) and the fluorescence produces a peak at 340 nm. As the protein is unfolded the peak wavelength of the fluorescence reduces in intensity.

2.5.2 Chemical Denaturation

2.5.2.1 Sample Preparation

A stock solution of 8 M Guanidinium Hydrochloride (GuHCl) was made up in ddH_2O and the accurate concentration was measured using a refractometer. A Hamilton Microlab autotitrator was fitted with two 1 ml Hamilton syringes and used to dispense ddH_2O and GuHCl (800 μl total volume) into 68 tubes varying the concentration of GuHCl linearly from 0-8 M. To each 800 μl tube of denaturant mixture, 100 μl of concentrated protein stock was added. The stock solution was 9 times the desired final concentration of protein, typically 45 μM. 1 M phosphate buffer (or 1 M MOPS depending on the protein buffer) was added to adjust the buffer to a final concentration of 50 mM. The protein samples were incubated at 10 ^{o}C for a minimum of 30 minutes to obtain equilibrium. The final protein concentration was 1-5 μM and buffer conditions were 50 mM phosphate pH 7 unless otherwise stated.

2.5.2.2 Chemical Denaturation followed by CD

An Applied Photophysics Chriascan spectrophotometer with a thermostated cuvette holder was used to collect data. The temperature used was 10 ^{o}C unless otherwise stated. A 5 mm pathlength cuvette was used for all experiments. The spectrum of each sample was recorded between 210-230 nm at an interval of 0.5 nm, scanning at 1 s per point. The ellipticity at 222 nm was plotted as a function of the denaturant concentration.

2.5.2.3 Chemical Denaturation followed by Fluorescence

Fluorescence was also used to follow the chemical denaturation. The sample preparation was the same as described in Section 2.5.2.1. A 5 mm pathlength cuvette was used. A Hitachi Fluostar F-2500 FL Fluorimeter with a thermostated cuvette

holder was used for all measurements. The samples were excited at 280 nm and the emission spectrum recorded between 320-360 nm at a scan speed of 300 nm/min. The Fluorescence at 340 nm was plotted as a function of the denaturant concentration.

2.5.3 Thermal Denaturation

2.5.3.1 Thermal Denaturation followed by CD

The protein sample, at a concentration of 5 μM, was prepared in 50 mM phosphate pH 7. A 5 mm pathlength cuvette was used. The temperature was ramped from 20-90 ^{o}C at intervals of 1 ^{o}C, with 30-60 s equilibration time at each temperature. The ellipticity was recorded at 222 nm and plotted against the temperature.

2.6 2-Dimensional Nuclear Magnetic Resonance (2D NMR)

The proteins: CTPR2, CTPR3, CTPRy2 and CTPRy3 were expressed and purified as N^{15} labelled samples (as described in Section 2.3.7.2)

2.6.1 Preparation of Protein Samples

N^{15}labelled protein samples were dialysed into 50 mM phosphate with 150 mM NaCl at pH 6.8. The samples were concentrated to 400-800 μM. For native state measurements: 10% (v/v) of D_2O was added. For partially denatured samples: concentrated GuHCl (8 M) was added to give the desired final concentration of denaturant, 10% (v/v) of D_2O was also added. CTPR2 and CTPRy2 were measured with 0.8 and 1.6 M GuHCl. CTPR3 and CTPRy3 were measured with 0.8, 1.6 and 2 M GuHCl. All samples were centrifuged at 14.1 krpm for 10 minutes to remove particulates prior to pipetting into an NMR tube.

2.6.2 Data Acquisition

Data was collected at 25 °C using a Bruker AV600 600 MHz spectrophotometer and processed using TopSpin 2.1 patch level 6. 2D 1H-15N HSQC spectra were acquired using Echo/Antiecho-TPPI gradient selection with decoupling during acquisition, using water flip-back pulse with gradients in reverse-inept. Acquisition parameters as follows: spectral widths, 15N, 60 ppm, 1H, 14 ppm; 2048 x 128 complex points acquired; acquisition time, 0.12 s; relaxation delay, 1 s.

2.6.3 Cross-peak Assignments

Spectra were analysed in TopSpin 2.1 patch level 6 Assignment of the cross-peaks was achieved by comparing the spectra of CTPR3 with published spectra for that construct.

2.6.4 Comparing data sets using Chemical Shift Perturbations

Chemical shift perturbations (CSP) values are used to quantify the changes observed between data sets for individual 2D NMR crosspeaks. For each crosspeak, the change in chemical shift for both the N^{15} and the H^1 spectra is calculated. These values are used in Equation 4.8 to calculate the CSP for each crosspeak.

$$CSP = \sqrt{\frac{\left(\frac{\delta \mathrm{N}}{5}\right)^2 + \left(\frac{\delta \mathrm{H}}{2}\right)^2}{2}} \tag{2.3}$$

δN is the change in the chemical shift between the N^{15} spectra. δH is the change in the chemical shift between the H^1 spectra.

2.7 Triggered Polymerisation of His-Xa-CTPR3ΔS-MESNa Proteins

The fusion protein His-Xa-CTPR3ΔS-MESNa was prepared as described in Section 2.3.8. Cleavage from the Chitin resin generates a thioester at the C-terminus. Factor Xa cleavage of the His-tag leaves a cysteine at the N-terminus. The cleavage of the His-tag triggers polymerisation of the monomers through Native Chemical Ligation (NCL) between the cysteine at the N-terminus and a thioester moiety at the C-terminus. The NCL reaction does not occur until the His-tag has been cleaved.

A standard reaction mixture contained 100μM protein in 50 mM phosphate pH 7 with 150 mM NaCl and 100 mM MESNa. However, various buffers conditions were trialled, including: 50 mM phosphate (pH 5, 6 and 7) with and without 150 mM NaCl; 50 mM Tris (pH 7 and 8) with and without NaCl. MESNa was included at a concentration of between 100-500 mM. The concentration of protein was varied from 50-500 μM. The reaction was performed at room temperature, with or without shaking.

2.7.1 Analysis of the Polymerisation mixture

The polymerisation reaction was followed by taking samples of the reaction mixture and analysing these samples using: SDS-PAGE, Transmission Electron Microscopy (TEM) and Size Exclusion Chromatography.

2.7.1.1 SDS-PAGE

Samples of the polymerisation mixture were analysed using SDS-PAGE. The same method was used as described in Section 2.4.2.1.

2.7.1.2 Transmission Electron Microscopy (TEM) of CTPR fibres

Time points of the polymerisation were analysed using TEM. The microscope used was a Jeol 1100. Samples were diluted 1:100 with ddH_2O. Various types of Carbon Formvar grids were trialled but typically a 400 mesh was used. Grids were glow discharged before use, to make them more hydrophilic to increase the binding of the sample. 4 μl of sample was dropped onto the grid, incubated for 1 minute and the excess removed using filter paper. 4 μl of Uranyl acetate stain (2 % w/v in ddH_2O) was added to the grid and incubated for 1 minute and the excess removed with filter paper, this step was repeated. The grid was allowed to fully dry before observation.

2.7.1.3 Size Exclusion Chromatography

A 10/300 Superdex 75pg (GE Healthcare) column (equilibrated with PBS pH 7) was used. 100 μl time point samples (at 50 μM) were loaded using a sample loop. The flow-rate was set at 0.5 ml/min and the elution was followed using UV at 280 nm.

2.8 AB-photolinker Ligation and Isomerisation

2.8.1 Ligation of the AB-photolinker onto the CTPR3 scaffold

The protein was dialysed into 50 mM Tris pH 8, 2 mM TCEP. A concentrated stock solution of 5 mM AB-photolinker in ddH_2O was prepared immediately before use. In total 500 nmol of AB-photolinker was reacted with 90 nmol of protein (approximately a 5 fold excess of linker). The linker was added to the 900 μl of protein (100 μM) in 3 aliquots (30, 30 and 40 μl) with an incubation of 20 minutes between each addition. The reaction was performed in the dark, at room temperature, with shaking, this was to ensure the linker was in the trans conformation. The linker was added in 3 aliquots, to reduce the likelihood of intermolecular cross-linking.

2.8.1.1 Assaying Ligation of the AB-photolinker using the Ellman's Reagent Test for free thiols

A solution of Ellman's reagent was prepared before each experiment by dissolving 4 mg of 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) in 3 ml of reaction buffer (100 mM phosphate with 1 M EDTA at pH 8). Each reaction contained 940 μl reaction buffer and 10 μl of protein sample. 50 μl of Ellman's reagent was added shortly before UV absorption spectra were collected between 300-600 nm. The number of free thiols can be calculated by using the Molar Extinction Co-efficient of TNB (14 150 $M^{-1}cm^{-1}$) and the Beer-Lambert law, Equation 2.1.

2.8.2 Isomerisation of the AB-photolinked-CTPR3

The isomerisation was accomplished using a 6 Watt 365 nm UV lamp. The protein sample was exposed to the light for 10-15 minutes.

2.8.2.1 Characterisation of Isomerisation of AB-photolinked-CTPR3 by UV Absorbance

UV absorbance wavelength scans were performed on a U-3010 Spectrophotometer scanning between 200-600 nm at room temperature. The protein concentration was between 15-25 μM in 50 mM Tris pH 8. Samples were measured in a 10 mm pathlength cuvette. The trans isomer absorbs at 350 nm, thus the isomerisation reaction was monitored by observing the change in absorbance at this wavelength.

2.8.2.2 Characterisation of Isomerisation of AB-photolinked-CTPR3 by CD

CD spectra were collected on an Applied Photophysics Chriascan spectrophotometer. A 5 mm pathlength cuvette was used in a thermostated cuvette holder held at 10 ^{o}C. Samples were diluted to a concentration of either 5 μM or 10 μM. The sample buffer was 50 mM Tris at pH 7 or pH 8. All CD scans were performed between 200-300 nm. An initial scan was performed in order to measure helical content of the fully folded protein. The sample was then isomerised using UV light at 365 nm as described in Section 2.8.2 and another wavelength scan performed. The α-Helical content of the protein was analysed by observing the Ellipticity at 222 nm.

2.9 Analysis of Equilibrium unfolding data

2.9.1 Two-state Analysis of Chemical Denaturation Unfolding

The simplest model to explain the thermodynamics of the unfolding of a protein is the two-state model. This states that the protein undergoes a single reversible transition from the folded to the unfolded state. The proteins studied during this thesis all displayed a single transition and were all fit according to the two-state model. Two-state unfolding is described as:

$$N \overset{K_D}{\longleftrightarrow} D \tag{2.4}$$

Where N is the folded protein and D is the unfolded protein. The equilibrium constant for the transition from folded to unfolded can be described as:

$$K_D = \frac{[D]}{[N]} \tag{2.5}$$

The equilibrium constant K_D is related to the free energy of unfolding ΔG_{D-N} via the following equation:

$$\Delta G_{D-N} = -RTlnK_D \tag{2.6}$$

The rate constant for unfolding can also be described as a function of the spectroscopic signal, λ_{obs}:

$$K_D = \frac{\lambda_N - \lambda_{obs}}{\lambda_{obs} - \lambda_D} \tag{2.7}$$

By rearranging Equation 2.6 for K_D we can substitute this into Equation 2.7 to give:

$$\frac{\lambda_N - \lambda_{obs}}{\lambda_{obs} - \lambda_D} = exp^{\left(-\frac{\Delta G_{D-N}}{RT}\right)} \tag{2.8}$$

When rearranged this gives:

$$\lambda_{obs} = \frac{\lambda_N + \lambda_D exp^{-\frac{\Delta G_{D-N}}{RT}}}{1 + exp^{-\frac{\Delta G_{D-N}}{RT}}} \tag{2.9}$$

As λ_D and λ_N are assumed to be linearly dependent on the denaturant concentration $[D]$ ($\lambda_N = \alpha_N + \beta_N[D]$ and $\lambda_D = \alpha_D + \beta_D[D]$):

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[D]) + (\alpha_D + \beta_D[D])\, exp^{-\frac{\Delta G_{D-N}}{RT}}}{1 + exp^{-\frac{\Delta G_{D-N}}{RT}}} \tag{2.10}$$

Here α_N and α_D are the y-intercepts and β_N and β_D are the gradients of the folded and unfolded baselines respectively.

Denaturant concentration varies linearly with ΔG, this is an empirical relationship [103]. The ΔG of unfolding can be expressed by Equation 2.11.

$$\Delta G^{D}_{D-N} = \Delta G^{H_2O}_{D-N} - m[D] \tag{2.11}$$

The gradient through the transition between folded and unfolded is defined as m. The m value is a constant of proportionality related to the change in solvent accessible surface area (SASA) of the protein as it transitions between the native and unfolded state. Equation 2.11 can be substituted into Equation 2.10 to give Equation 2.12:

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[D]) + (\alpha_D + \beta_D[D])\, exp^{-\frac{\Delta G^{H_2O}_{D-N} - m[D]}{RT}}}{1 + exp^{-\frac{\Delta G^{H_2O}_{D-N} - m[D]}{RT}}} \tag{2.12}$$

At the midpoint of denaturation ($[D]_{50\%}$) in a two-state approximation, half of the population of the protein molecules are folded and half are unfolded. Thus Equation 2.11 can be rearranged to:

$$\Delta G^{H_2O}_{D} = m[D]_{50\%} \tag{2.13}$$

Equations 4.3 can be substituted in to Equation 2.11 to give:

$$\Delta G^{D}_{D-N} = m[D]_{50\%} - [D] \tag{2.14}$$

Equation 2.14 can be substituted into Equation 2.12 to give:

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[D]) + (\alpha_D + \beta_D[D])\, exp^{\left(-\frac{m([D]-[D]_{50\%})}{RT}\right)}}{1 + exp^{\left(-\frac{m([D]-[D]_{50\%})}{RT}\right)}} \tag{2.15}$$

To calculate $[D]_{50\%}$ and m, the spectroscopic signal (λ_{obs}) was plotted against the denaturant concentration in Kaleidagraph version 4 (Synergy Software). The data was fitted to equation 4.2 using linear regression analysis in Kaleidagraph. Once $[D]_{50\%}$ and m are known, Equation 4.3 is used to calculate ΔG_D.

2.9.2 Normalisation of the Chemical Denaturation Equilibrium Unfolding Data

The Fluorescence or Circular Dichroism spectroscopic signal of a CTPR varies with the number of repeats. When comparing CTPR2 and CTPR3, CTPR2 contains less tryptophans or tyrosines and produces a lower Fluorescence signal. CTPR2 has less α-helices and produces a lower CD signal.

$$\lambda_{norm} = \frac{\lambda_{222} - \alpha_N}{\alpha_D - \alpha_N} \tag{2.16}$$

The values for α_D and α_N are the y-intercept values of the unfolded and folded baselines. This equation allows for the data to retain the sloping baselines.

2.9.3 Analysis of Thermal Unfolding

The observed signal for thermal denaturations varies with temperature according to Equation 2.9. As λ_D and λ_N are assumed to be linearly dependent on the temperature T, this gives the following equation:

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N T) + (\alpha_D + \beta_D T)\, exp^{-\frac{\Delta G_{D-N}}{RT}}}{1 + exp^{-\frac{\Delta G_{D-N}}{RT}}} \tag{2.17}$$

The free energy of unfolding ΔG_{D-N} can be expressed using the Gibbs-Helmholtz relationship, shown in Equation 2.18.

$$\Delta G_{D-N} = \Delta H_m \left(1 - \frac{T}{Tm}\right) + \Delta C_p \left[T - T_m - \left(T.ln\frac{T}{T_m}\right)\right] \tag{2.18}$$

Where ΔG_{D-N} is the enthalpy of unfolding, ΔH_m is the enthalpy of denaturation at the midpoint, ΔC_p is the change in heat capacity of denaturation and T_m is the midpoint of the denaturation. ΔC_p is unique to each protein and can be determined through calorimetry. Values for ΔC_p were not determined as part of this thesis, therefore ΔC_p was taken to be zero. This changes Equation 2.18 into:

$$\Delta G_{D-N} = \Delta H_m \left(1 - \frac{T}{Tm}\right) \tag{2.19}$$

When Equation 2.19 is substituted into Equation 2.17:

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[T]) + (\alpha_D + \beta_D[T])\, exp^{-\Delta H_m\left(\frac{1-\frac{T}{T_m}}{RT}\right)}}{1 + exp^{-\Delta H_m\left(\frac{1-\frac{T}{T_m}}{RT}\right)}} \tag{2.20}$$

Thermal denaturation data was plotted with λ_{obs} against temperature in Kaleidagraph 4.0 and globally fit to Equation 2.20

2.9.4 Normalisation of the Thermal Denaturation Equilibrium Unfolding Data

The ellipticity of a CTPR varies with the number of repeats. When comparing CTPR2 and CTPR3, CTPR2 has less α-helices and produces a lower CD signal.

$$\lambda_{norm} = \frac{\lambda_{222} - \alpha_N}{\alpha_D - \alpha_N} \tag{2.21}$$

α_D is the λ_{222} at 90 °C and α_N is the λ_{222} at 20 °C. This equation allows for the data to retain the sloping baselines.

Chapter 3

Self-Assembly of CTPR Modules

In this chapter I will describe the self-assembly of CTPR3ΔS modules into fibrous nanostructures. The fibres are formed using Native Chemical Ligation (NCL) between monomers. NCL generates a peptide bond from the reaction between an N-terminal cysteine and a C-terminal thioester.

3.1 Experimental Approach

The formation of the CTPR fibres was monitored using SDS-PAGE and Analytical Size Exclusion Chromatography. The fibres were imaged using Transmission Electron Microscopy (TEM). The size and morphology was analysed from the TEM images.

3.2 Motivation for using the TPR motif to Create Nanostructures

A **T**etratrico**P**eptide **R**epeat (TPR) is a 34-residue helix-turn-helix motif, which when stacked together forms elongated domains [13]. Unlike globular proteins, TPRs do not rely on topologies stabilised by interactions of residues distant in primary sequence. Instead, their modular, non-globular structures are dominated by regularized interactions of residues close in primary sequence [13]. These distinctive features have made designing proteins that contain arrays of repeats composed of consensus residues extremely successful [13] [8]. Significantly for biomaterial synthesis, these designs have shown that the consensus proteins can be easily lengthened or shortened by the addition or removal of identical repeat motifs [8, 75, 104]. In particular, we and the Regan laboratory have recombinantly produced soluble, monomeric proteins with up to 20 consensus TPR modules [8], [9, 72, 105]. Strikingly, as you increase consensus TPR motifs within a protein, large elongated super-helical structures with identical inter and intra repeat interactions are produced [105]. Moreover, increasing the number of stacked repeats gives proteins that increase in stability [9,105]. Thus if an arbitrarily large structure comprised of consecutively arrayed linear repeat units could be produced, it should form helical filaments with a free-energy stability orders of magnitude greater than that of the soluble monomers. One method of producing larger superstructures would be through further recombinant expression of repeat proteins with greater repeat numbers within a single gene. However such a process is severely limited, not least in the size and yield of protein that can be produced. Therefore, we chose a strategy of bottom-up directed self-assembly. Specifically, the polymerisation of recombinantly produced, discretely folded, TPR protein domains. These monomers were based on the original designed CTPR3 protein (3 consensus TPR motifs of 34 amino acids each) [8]. CTPR3 was chosen as it is extremely easy to recombinantly synthesize [8], is highly stable (12.0 $\pm$ 0.7 kcalmol^{-1}, at pH

7 [78]), is structurally very rigid 20 and contains a minimal 3 TPR motif unit that is used ubiquitously throughout nature as a peptide binding motif [25]. These features are important as they provide abundant building blocks that remain folded under a range of conditions, whose rigidity prevents futile intra-protein cyclisation and present a viable route for future decoration and functionalisation.

3.3 Results

3.3.1 Developing CTPR3 into a Suitable Building Block for Polymerisation

The CTPRs were designed to be monomeric through the addition of a C-terminal 'capping' helix. This was added to aid stability and cap the hydrophobic core. The C-terminal helix in CTPR3 is equivalent to half a repeat and consequently blocks a potential dimerisation interface at the C-terminus, Figure 3.1.

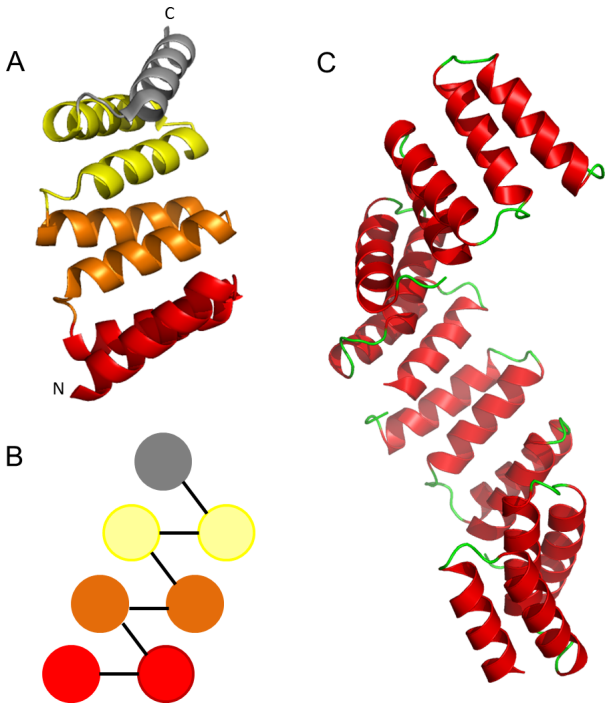

Figure 3.1: (A) Crystal structure of CTPR3 (1NA0) (B) Schematic diagram of CTPR3 with the helices represented as circles. (C) CTPR8 (2FO7) showing the super-helix that is formed when the number of repeats exceeds 7.

To make a monomer suitable for further polymerisation you need compatible interfaces and the first step towards redesigning CTPR3 was to remove the C-terminal helix. Removing this helix exposes a surface that is identical to that seen between repeats and is compatible with the interface at the N-terminus of another CTPR3

monomer, thus allowing specific head-to-tail association.

3.3.2 Cloning, Expression and Purification of CTPR3ΔS

This mutant that lacks the C-terminal helix is termed CTPR3ΔS and was created by inserting a stop codon at position 106 in the CTPR3 sequence (see Section 2.1, Table 2.1 for a full description). It was expressed and purified as described in Chapter 2.3.2 and gave high yields of approximately 10 mg/L of culture.

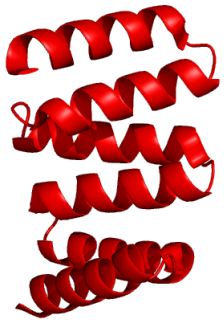

Figure 3.2: Structure of CTPR3ΔS, generated from the crystal structure of CTPR3 (1NA0)

3.3.3 Characterisation of the CTPR3ΔS Protein

Removal of the C-terminal helix from CTPR3 produces a protein that has compatible dimerisation interfaces at both the N and the C-terminus. Therefore it was assumed that removal of this helix would lead to specific head-to-tail association between CTPR3ΔS units. An increase in non-specific aggregation was also anticipated as when naturally occurring TPR proteins were expressed and purified, they displayed low solubility without the addition of an extra C-terminal helix [8]. To determine if these previous assumption are correct, CTPR3ΔS was analysed. Size Exclusion Chromatography was used to analyse if it was monomeric. Circular Dichroism was used to investigate the structure. Thermal and GuHCl denaturations (followed by CD) were carried out to assess the stability.

3.3.3.1 Size-Exclusion Chromatography of CTPR3ΔS

CTPR3ΔS was analysed using size exclusion chromatography to determine if it is monomeric. A 100 μl sample at a concentration of 50 μM was loaded onto a GE Healthcare Superdex 75pg 300/10 column (equilibrated with 50 mM phosphate pH 7, 150 mM NaCl). The flow rate was set at 0.5 ml/minute. The elution of the protein was monitored using UV absorbance at 280 nm. The chromatogram for CTPR3ΔS was compared with that of CTPR3 (Figure 3.3) because CTPR3 has been confirmed as monomeric [8]. Both proteins show a single peak (CTPR3: 10.6 ml and at CTPR3ΔS: 10.9 ml), the variation in retention volume is because CTPR3 is a larger protein due to having an extra helix. CTPR3ΔS was judged to be monomeric up to a concentration of 50 μM. The removal of the C-terminal helix didn't result in an increase in non-specific aggregation, and CTPR3ΔS was monomeric as judged by size exclusion chromatography under the conditions that were used here, see Figure 3.3.

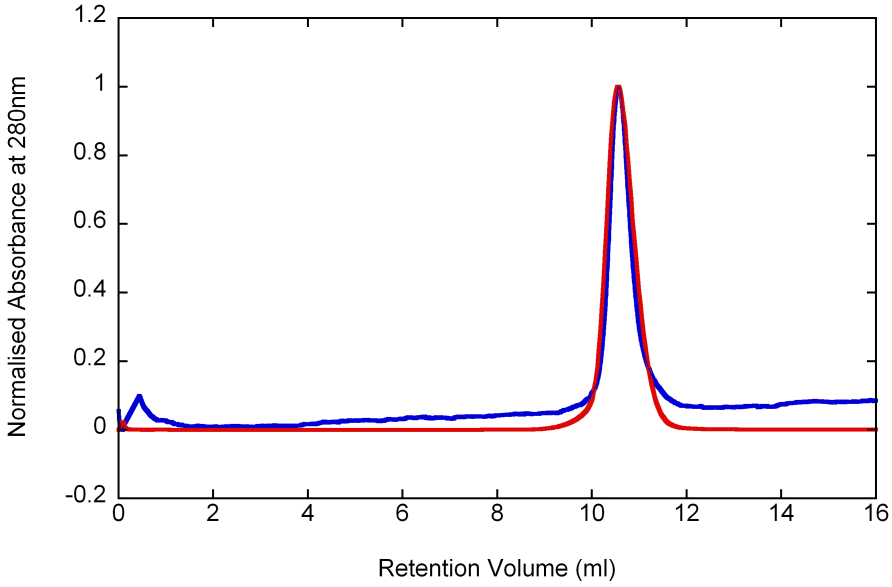

Figure 3.3: Analytical size exclusion chromatograms of CTPR3 (blue) and CTPR3ΔS (red). Conditions were: 50 mM phosphate pH 7, 150 mM NaCl, buffer flow rate was set at 0.5 ml/minute. The column used was a GE Healthcare Superdex 75pg 300/10.The elution of the protein was monitored using the absorbance at 280 nm and the retention on the column measured in ml. The increased retention time for CTPR3ΔS is due to the removal of the C-terminal helix, this makes it a smaller construct than CTPR3.

3.3.3.2 Analysing the Secondary Structure of CTPR3ΔS using Circular Dichroism

A CD wavelength scan was performed on CTPR3ΔS to assess the secondary structure, Figure 3.4. The wavelength scan was performed on a 5 μM sample, dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. The scan showed a single negative peak at 222 nm, indicating that the protein is highly α-helical. This is what was expected given the highly α-helical content of other CTPR proteins. This confirms that CTPR3ΔS displays the characteristic fold of the CTPRs.

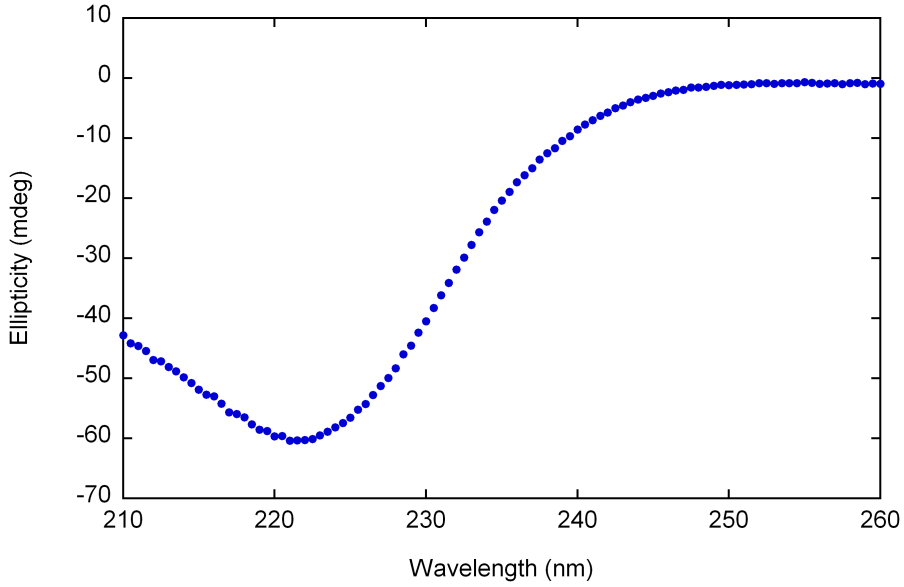

Figure 3.4: Far UV wavelength scan of CTPR3ΔS at 5 μM dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. The scan showed a single negative peak at 222 nm, indicating that the protein is highly α-helical.

3.3.3.3 Assessing the Stability of CTPR3ΔS Through Equilibrium Unfolding

The stability of CTPR3ΔS was assessed via thermal and GuHCl denaturations followed using Circular Dichroism at 222 nm (Figure 3.5). Through comparison with data for CTPR3 and CTPR2 we can judge the effect of removing the C-terminal helix on the stability of the protein (Figure 3.5).

3.3.3.3.1 Chemical Denaturation of CTPR3ΔS

CTPR3ΔS underwent a single reversible transition from the folded to the unfolded state. The data for CTPR3ΔS can be qualitatively compared to data for CTPR2 and CTPR3 by fitting to a two-state two-sloping baseline equation 4.2. The mid-points of the fit of the data can be used to compare the stability of the different proteins. These were: CTPR2 2.6 M ± 0.01 M; CTPR3ΔS 3.2 M ± 0.01 M and CTPR3 3.4 M ± 0.1 M (errors are calculated from the fit of the equilibrium unfolding curve).

Thus removing the C-terminal helix reduces the stability of the protein.

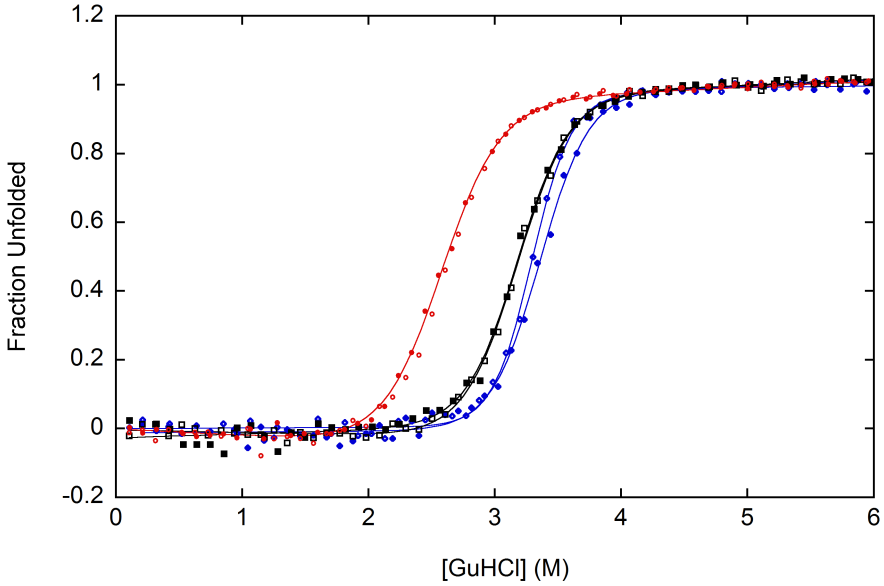

Figure 3.5: Normalised chemical denaturation curves for CTPR2 (red circles), CTPR3 (Blue diamonds) and CTPR3ΔS (black squares), there are 2 data sets for each construct. All the proteins undergo a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baseline equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation in ellipticity that arises from the different number of helices for each construct. Data was fit using linear regression analysis in Kaleidagraph 4.0. The data shows that the stability increases with the size of the protein, as judged by the mid-point of the denaturation. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

3.3.3.3.2 Thermal Denaturation of CTPR3ΔS

CTPR3ΔS underwent a single reversible transition from the folded to the unfolded state. Reversibility was confirmed by comparing CD wavelength scans before and after the denaturation, Figure 3.6A. The variance of the ellipticity at 222 nm between the wavelength scans was within 5%. The data for CTPR3ΔS can be qualitatively compared to data for CTPR2 and CTPR3 by fitting to a two-state two-sloping baseline equation 2.20. The mid-points of the fit of the data can be used to compare the stability of the different proteins.

CTPR3ΔS has the same mid-point (T_m) as CTPR2 (69 °C ± 0.2 °C). This is in contradiction to the GuHCl denaturation data, where the mid-point of CTPR3ΔS is between that of CTPR2 and CTPR3. The explanation for this is that the thermal transition is not a two-state mechanism and the protein unfolds partially before the transition, Figure 3.6B.

Although the C-terminal helix was added to reduce non-specific aggregation and block specific oligomerisation, CTPR3ΔS did not display either of these effects. CTPR3ΔS is stable and monomeric at pH 7 and didn't undergo any non-specific aggregation under the conditions examined here. Although CTPR3ΔS did not spontaneously oligomerise, it is very stable and doesn't aggregate thus making it a suitable building block for creating nanostructures.

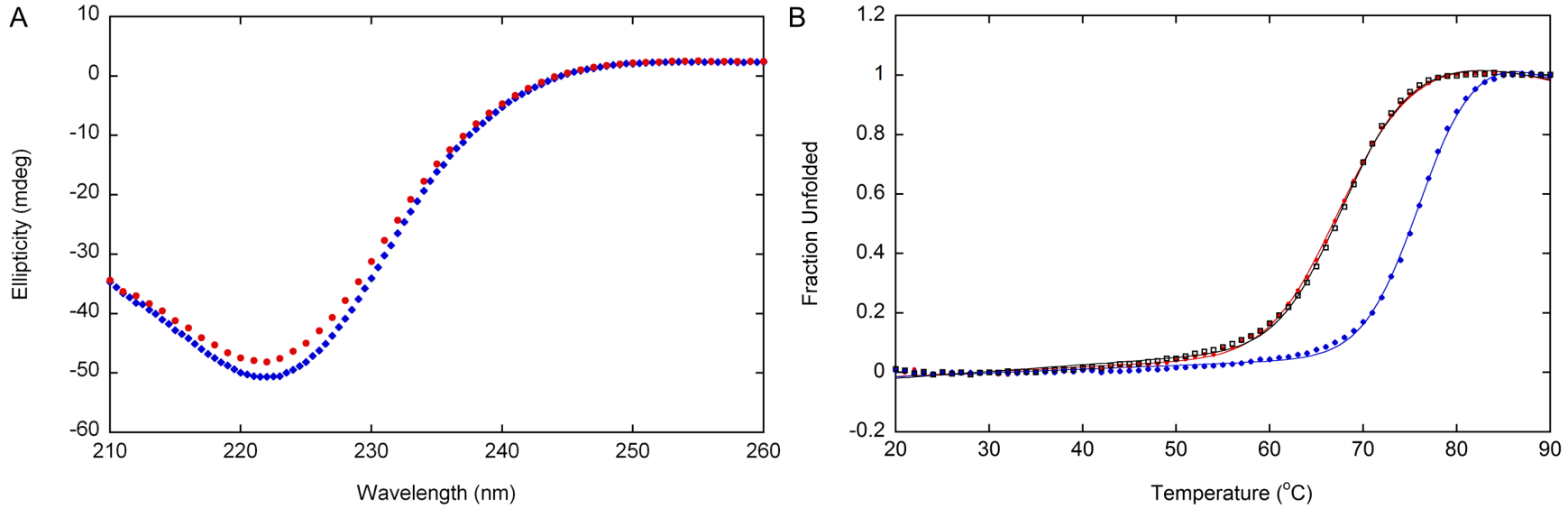

Figure 3.6: (A) Circular Dichroism wavelength scans of the CTPR3ΔS protein before (blue diamonds) and after the thermal melt (red circles). The protein concentration is 3 μM in 50 mM phosphate pH 7, the temperature of sample was maintained at 20 °C with a thermostated cuvette holder. These data shows that the thermal unfolding is reversible as the ellipticity at 222 nm returns to within 5% of the initial value. (B) Normalised thermal melt data, for CTPR2 (red circles), CTPR3 (Blue diamonds) and CTPR3ΔS (black squares). All the proteins undergo a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baseline equation. However, to allow for easier comparison the data has been normalised (using the equation described in Section 2.9.4)to account for the variation in ellipticity that arises from the different number of helices for each construct. Data was fit using linear regression analysis in Kaleidagraph 4.0 (Section 2.9). The data shows that CTPR3ΔS is as thermally stable as CTPR2. This result does not follow the same trend seen for GuHCl denaturation, where an increase in size results in an increase in stability. This is most likely due to the CTPR3ΔS undergoing a multistate transition. A multistate transition would mean that the CTPR3ΔS partially unfolds before the main transition, which would lower the mid-point relative to CTPR2. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. The data was collected on a Photophysics Chirascan using a 5 mm pathlength cuvette. The temperature of the sample was ramped from 20 °C to 90 °C in 1 °C increments, with 1 minute of incubation at each temperature. The ellipticity at 222 nm was recorded for each temperature.

3.3.4 Using a Bi-functional Chemical Crosslinker to test for Associations Between CTPR3ΔS Proteins

Although CTPR3ΔS was monomeric as judged by Size Exclusion Chromatography, it's possible that the protein is weakly associating, which won't be observed using Size Exclusion Chromatography. To examine if CTPR3ΔS does form weak or transient interactions, a bi-functional crosslinker was selected. Cysteine residues are the most common target for crosslinking reagents. However because CTPRs do not contain any cysteines, the lysine reactive linker: Suberic acid bis(3-sulfo-N-hydroxysuccinimide ester) was chosen (Figure 3.7). CTPR proteins contain 2 lysine residues per repeat and these are both solvent accessible according to the PDB data, Figure 3.8.

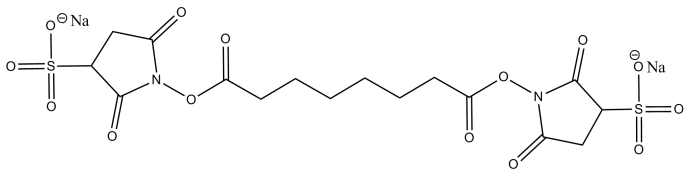

Figure 3.7: Chemical structure of the bi-functional linker, Suberic acid bis(3-sulfo-N-hydroxysuccinimide ester) sodium salt. The sulfo groups allow the linker to be dissolved in water.

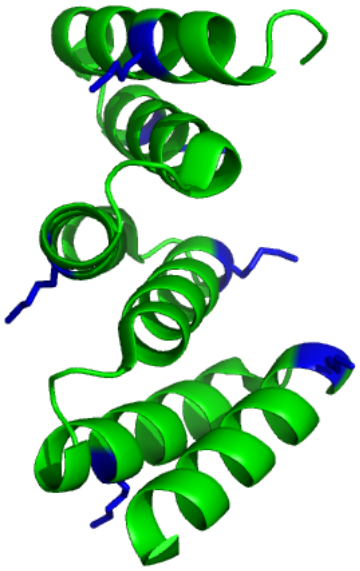

Figure 3.8: Structure of CTPR3ΔS (modified 1NA0) with the lysine residues highlighted in blue.

Crosslinks between protein molecules are only formed when the protein molecules are near in space. This allows us to observe any weak or transient interactions that are occurring between the proteins. If the CTPR3ΔS proteins form constructive head-to-tail associations then dimers and trimers will form.

3.3.4.1 Crosslinker Reaction Conditions

The protein and linker were reacted in a 1:50 mole ratio to give final concentrations of 90 μM protein and 4.5 mM BS3. The reaction was performed at 25 °C and 50 °C in 20 mM phosphate pH 7. CTPR3 was reacted with the linker in a separate vessel to serve as a control. The lysine residues are located throughout the structure of CTPR3 and CTPR3ΔS, therefore it is possible for non-specific dimers to be formed on reaction of the crosslinker with the protein. CTPR3 is prohibited from forming specific head-to-tail associations because of the C-terminal helix but will still form non-specific dimers. The reaction was performed at 50 °C to determine if the C-terminal helix of CTPR3 could be specifically unfolded and yield a similar result as CTPR3ΔS. Samples of the reaction were analysed using SDS-PAGE and are shown

in Figure 3.9.

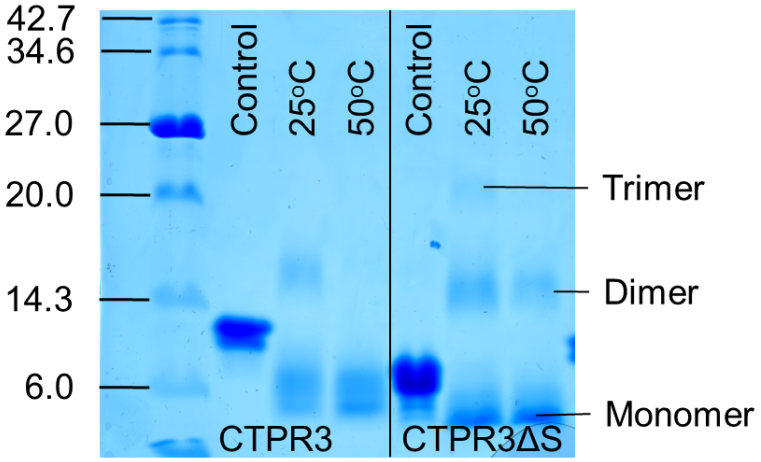

Figure 3.9: 18% SDS-PAGE gel of samples crosslinked with BS3. The protein and linker were reacted in a 1:50 mole ratio to give final concentrations of 90 μM protein and 4.5 mM BS3. Dimers are seen for both CTPR3 and CTPR3ΔS but trimers are only seen for CTPR3ΔS. The reaction was performed at different temperatures to assess if temperature can destabilise the C-terminal helix in CTPR3. The control sample on the gel is a sample of the protein without any crosslinker added, the apparent difference in size is caused by single attachment of the crosslinker to the protein, this increases the negative charge thus increasing the speed.

When this linker was reacted with CTPR3 and CTPR3ΔS there were some dimers seen for both proteins but a higher number was seen with CTPR3ΔS. There were some trimers formed from CTPR3ΔS but not for CTPR3. This indicates that the CTPR3ΔS is associating constructively to form trimers. When the reaction was performed at 50 °C, there was a reduction seen in the number of higher order oligomers. The reaction between the crosslinker and the free amines competes with hydrolysis. Hydrolysis is the dominant effect seen at 50 °C resulting in less crosslinks. This shows that there are some transient associations formed between CTPR3ΔS monomers. However, this would not be a suitable system for creating nanostructures as there are many non-specific crosslinks formed and the reaction is very susceptible to hydrolysis. To ensure that only constructive oligomers are formed, a way of driving specific head-to-tail polymerisation needed to be devised.

3.3.5 Redesigning CTPR3ΔS so that it can be triggered to polymerise

The next step towards inducing CTPR3ΔS polymerise is to engineer orthogonal chemistry into both ends to drive polymerisation in the desired head-to-tail direction. We wanted to use a semi-synthetic approach that could be genetically encoded. For this reason Native Chemical Ligation was chosen. NCL has several distinct advantages over other chemistries as it requires no protecting groups, can be genetically encoded, works at neutral pH in aqueous buffer, kinetics are productive within a temperature range in which the proteins are natively folded, it requires no catalyst and results in the formation of a native peptide bond, such that the site of ligation is indistinguishable from the rest of the polypeptide chain.

3.3.5.1 Native Chemical Ligation

Native Chemical Ligation (NCL) is used to link proteins by forming a peptide bond between proteins. The pre-requisites for inter-molecular NCL are: a cysteine residue at the N-terminus and a thioesterified C-terminus. Figure 3.10 shows the ligation of two proteins via NCL. In step 1 the thiol group of an N-terminal cysteine residue attacks the thioester at the C-terminus of another protein. Next, the amine group attacks the carbonyl giving an $S \rightarrow N$ acyl shift, that leaves a native peptide bond at this site. If CTPR3ΔS is redesigned to possess an N-terminal cysteine and a C-terminal thioester then NCL can occur between these monomers and result in polymerisation.

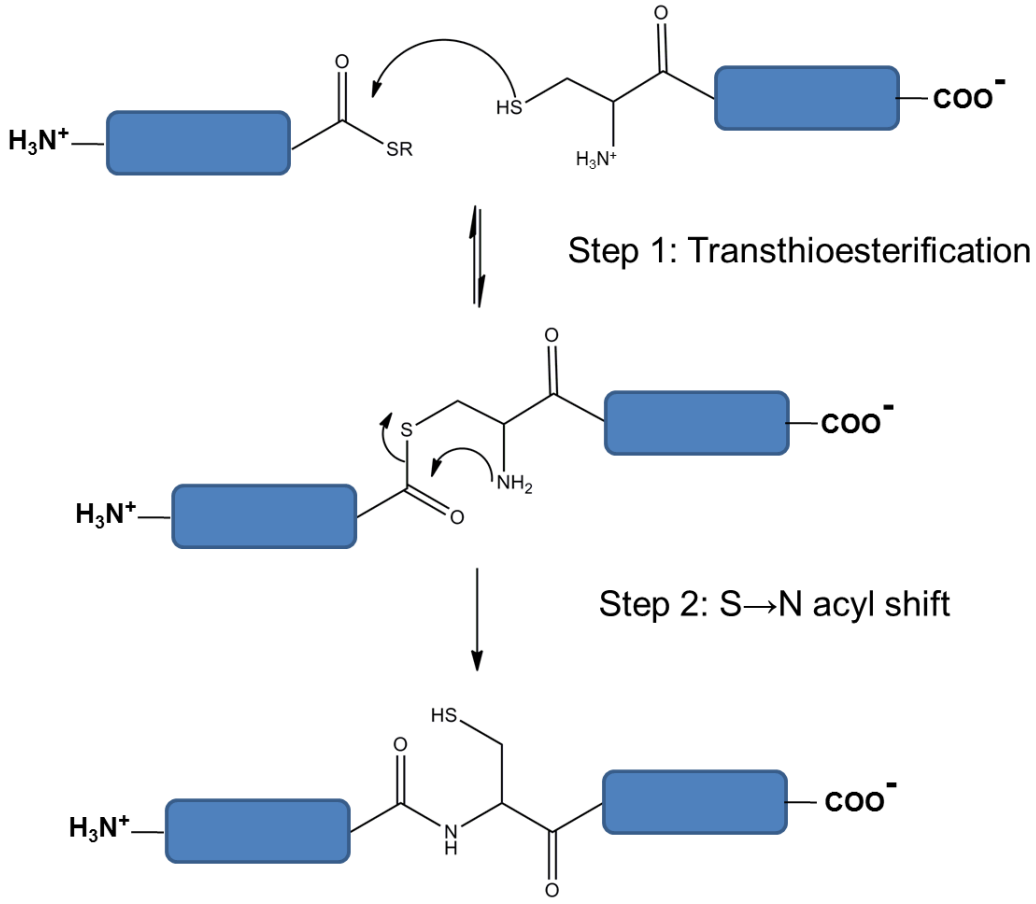

Figure 3.10: Reaction mechanism for a Native Chemical Ligation between two protein molecules (shown in blue) at pH 7. Step 1: The thiol group of a cysteine at the N-terminus attacks the carbonyl of the thioesterified C-terminus of another protein. This step is reversible and only proceeds to the next step when the R-group of the thioester is released. Step 2: The $S \rightarrow N$ acyl shift results in the formation of an amide bond.

3.3.5.2 IMPACT™Intein-mediated Protein Ligation kit

The IMPACT™kit from New England Biolabs is used to express proteins ready for use in a Native Chemical Ligation reaction. This system utilises the vectors pTWIN1 and pTWIN2. These vectors are used to express fusion proteins with intein domains. An intein is a self-cleaving protein domain and this system uses this feature to cleave the affinity tag as an alternative to using proteases. A gene can be ligated into the vector to produce fusion proteins with either an N-terminal intein, a C-terminal intein or both. A chitin binding domain (CBD) allows the fusion protein

to be purified using chitin resin. The mechanism for cleavage from the chitin resin depends on whether an N or C-terminal tag is being removed. The mechanism for cleavage of a C-terminal tag requires a thiol reducing agent, Figure 3.11A-C. The process for cleaving the fusion protein from the chitin resin starts with the $N \rightarrow S$ acyl shift (Figure 3.11A). The reducing agent, Mercaptoethanesulphonate Sodium salt (MESNa), is added and this attacks the carbonyl in the protein backbone (Figure 3.11B) and the protein is released from the resin with a thioester at the C-terminus (Figure 3.11C). The process for removing an N-terminal tag is achieved through a shift in pH, Figure 3.11D-E. The change in pH results in the amide group of an upstream asparagine residue attacking the backbone cleaving the protein here. The advantage of this mechanism of cleavage over that of a protease is that the choice of N-terminal residue isn't constrained to that required by the protease. A cysteine can be placed at the N-terminus, this is what is required for NCL.

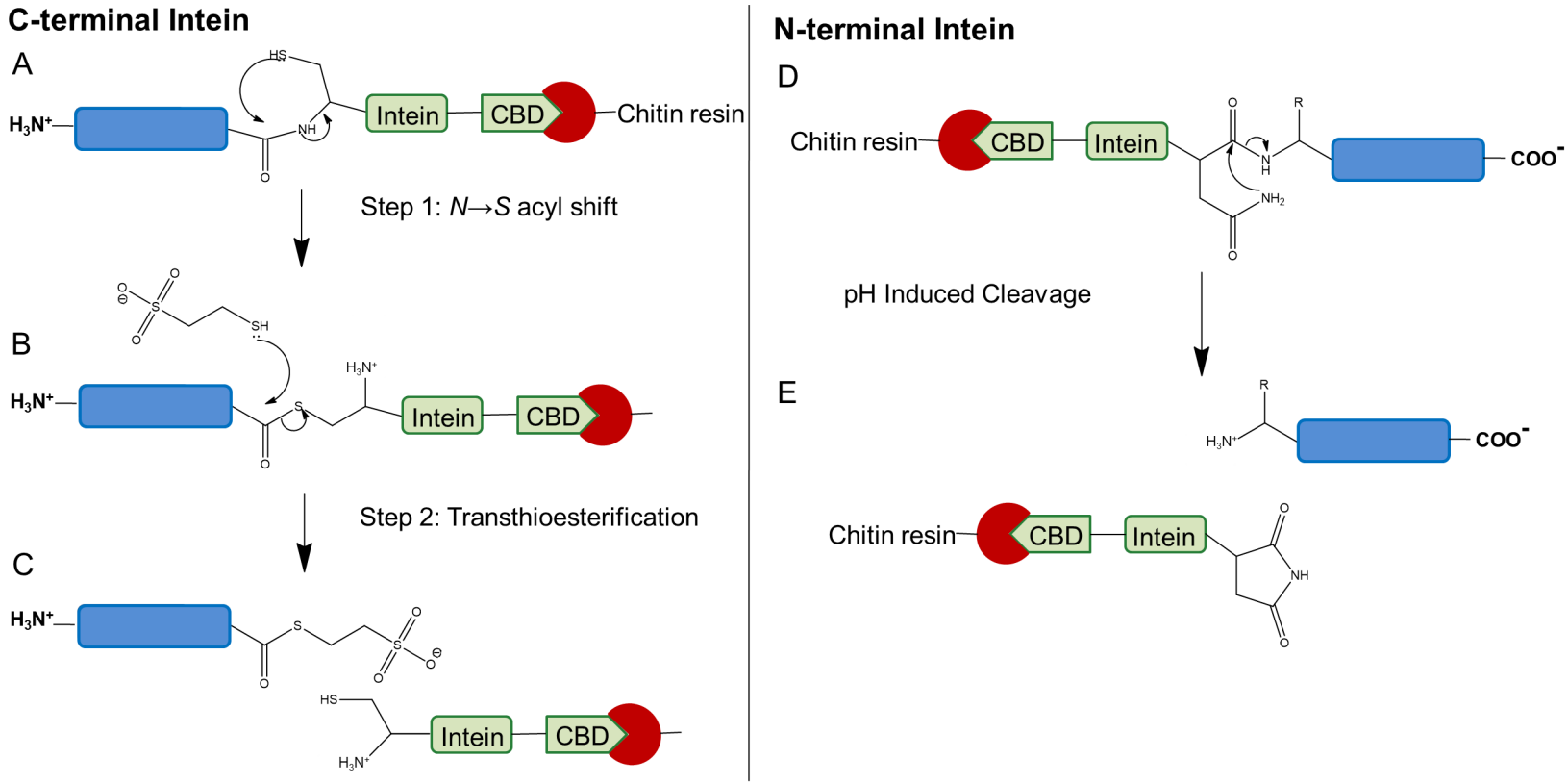

Figure 3.11: (A-C) Purification of a C-terminal intein fusion protein with a chitin binding domain (CBD). (A) The full length fusion protein, target protein (blue) and intein and CBD (green), is bound via the CBD to chitin resin (red). The protein undergoes a spontaneous $S \rightarrow N$ acyl shift this generates a thioester. (B) The thiol group of the reducing agent, Mercaptoethanesulphonate Sodium salt (MESNa), attacks the carbonyl group in a transthioesterification reaction. (C) The protein is released from the resin and now possess a thioester at the C-terminus. The intein sequence remains bound to the chitin resin. (D-E) Purification of an N-terminal intein fusion protein with a chitin binding domain (CBD). The full length fusion protein, target protein (blue) and intein and CBD (green). A change in pH results in the intramolecular attack of the amide group of asparagine, resulting in the cleavage of the amide bond. The N-terminal residue on the target protein can be any residue.

3.3.5.3 Developing CTPR3ΔS into a protein that can undergo Triggerable Cleavage

A cysteine residue was added to start of the CTPR3ΔS sequence and this was ligated into the pTWIN vector from the IMPACT™kit, so that the fusion protein that was expressed contained the N and C-terminal intein domains. This was so that MESNa induced cleavage introduced a thioester at the C-terminus and a change in pH would cleave the N-terminal tag.

However, when CTPR3ΔS was expressed as this fusion the protein underwent a significant amount of non-specific aggregation on cleavage of the N-terminal tag. The cleavage of the C-terminal intein was successful. Therefore, the construct was redesigned to retain the intein (Mxe GyrA [106]) and the chitin binding domain sequence at the C-terminus but the N-terminal intein was removed. Instead the N-terminal cysteine is generated through Factor Xa cleavage. Factor Xa was chosen as the protease to remove this N-terminal cap as it cuts after its recognition sequence, allowing a cysteine to be located as the very first residue at the N-terminus following cleavage. This gene was ligated into the pTrcHis-TOPO vector, this adds a His-tag to the N-terminus. This means that polymerisation is triggered by the addition of this protease. This strategy produced protein (A) in Figure 3.12. The reaction scheme in Figure 3.12 shows how the reactive monomer (C) is formed from the fusion protein (A).

The mechanism for polymerisation between activated CTPR3ΔS monomers is shown in Figure 3.12D-F. After the NCL reaction between the monomers the CTPR3ΔS units rearrange and dock so that the newly formed dimer adopts the characteristic TPR fold.

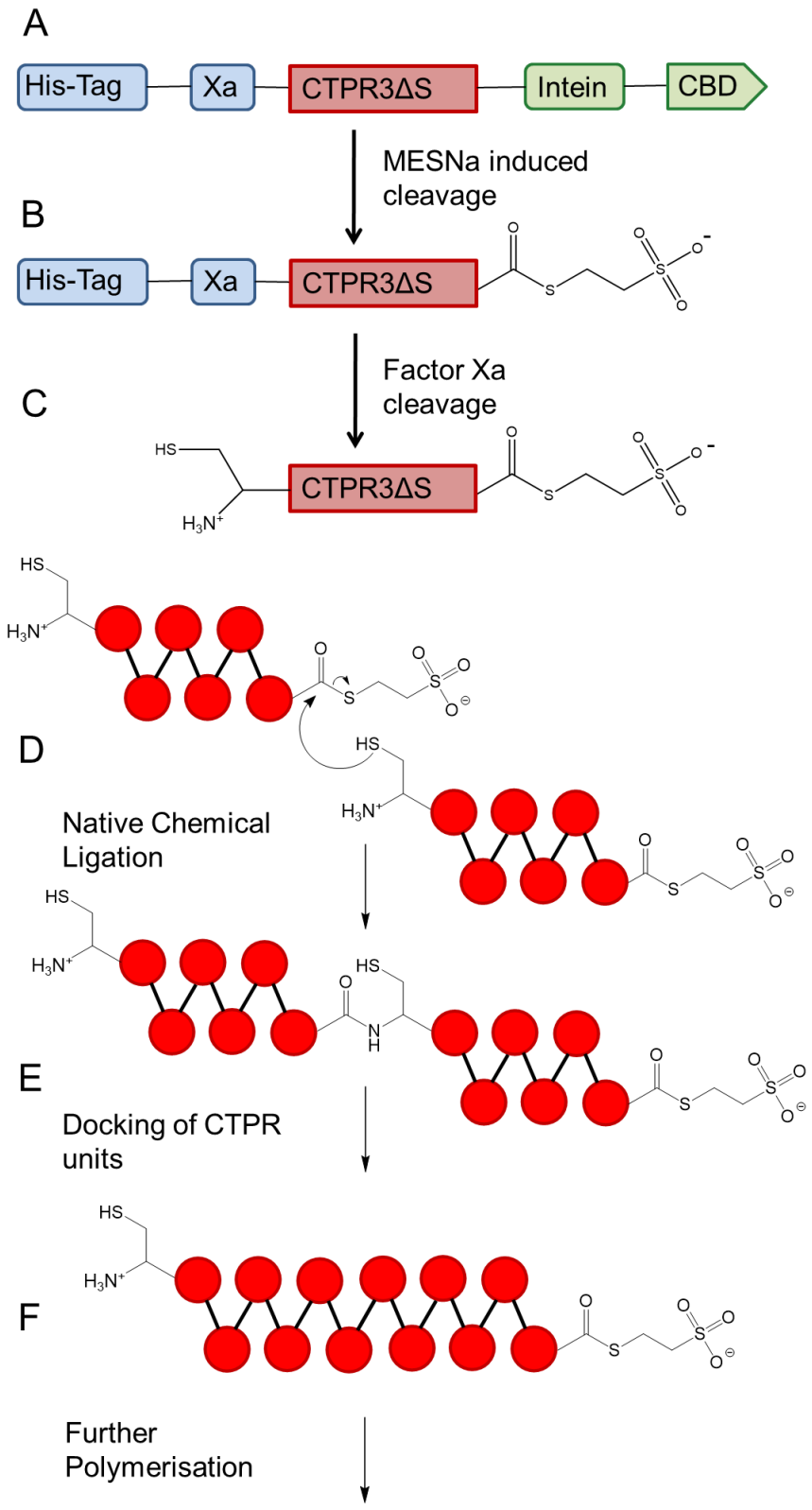

Figure 3.12: Reaction scheme showing how the reactive monomer is formed from the fusion protein (A). First, MESNa-mediated cleavage generates a thioester at the C-terminus (B), then Factor Xa cleavage leaves a cysteine at the N-terminus (C). (D) Shows the NCL reaction between 2 units of the active monomer to form a dimer (E). This dimer can rearrange and the CTPR3ΔS units will dock together (F), this dimer has reactive ends and can undergo further polymerisation.

3.3.6 Expression of the His-Xa-CTPR3ΔS-Intein-Chitin binding domain protein

The full length fusion protein (His-Xa-CTPR3ΔS-intein-Chitin binding domain) was expressed using the same protocol as described in Section 2.3.2. The cells were lysed under native conditions as described in Section 2.3.2.2.

3.3.7 Purification of the His-Xa-CTPR3ΔS-thioester Protein

The His-Xa-CTPR3ΔS-intein-CBD protein was bound to chitin resin using gravity flow. Cleavage is accomplished by incubating the resin bound protein overnight in a solution that contains 0.5 M of the reducing agent MercaptoEthaneSulfonateNa (MESNa). The His-Xa-CTPR3ΔS-thioester protein is released from the resin after undergoing intein-mediated cleavage, Figure 3.11. MESNa is required to generate the thioester at the C-terminus of His-Xa-CTPR3ΔS-thioester as the protein is cleaved from the intein-CDB-Chitin resin complex. The high specificity of the chitin binding domain allows for the purity of the His-Xa-CTPR3ΔS-thioester protein to be >95 %, see Figure 3.13. The intein-CBD largely remains bound to the resin but small amount is released along with the His-Xa-CTPR3ΔS-thioester protein. This is seen in the gel in Figure 3.13 at 27 kDa in lanes 5-9.

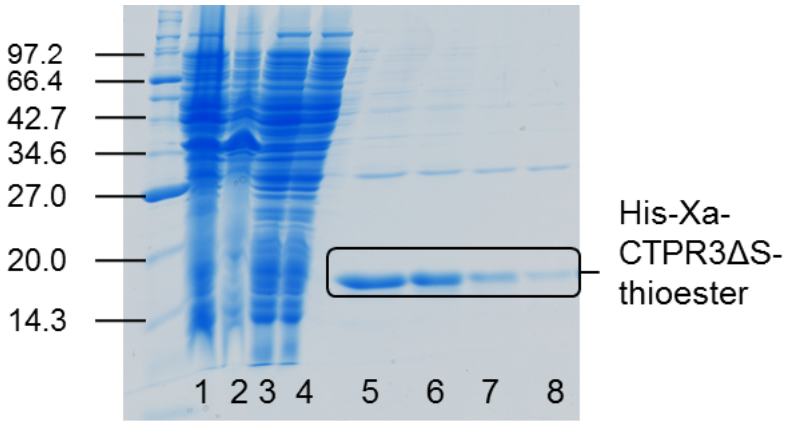

Figure 3.13: 18% SDS-PAGE gel of the purification of the intein fusion protein. The marker is in the left-hand Lane 1-4, form left to right are: lysed cells soluble fraction; lysed cells insoluble fraction; Wash of the soluble fraction through the column. Lanes 5-9 show the 4 elution fractions of protein (His-Xa-CTPR3ΔS-thioester at approximately 17 kDa) after it have been cleaved from the chitin column using MESNa. The band at 27 kDa that is present in lanes 5-9 is the intein-CBD protein, most of the concentration of this protein stays bound to the chitin resin but a small amount is released.

3.3.8 Native Chemical Ligation Reaction of the His-Xa-CTPR3ΔS-thioester Protein

The Native Chemical ligation reaction is triggered by the addition of Factor Xa to the His-Xa-CTPR3ΔS-thioester protein. This cleaves the His-tag from His-Xa-CTPR3ΔS-thioester to leave an N-terminal cysteine, Figure 3.12. The CTPR3ΔS-thioester protein then undergoes repeated NCL reactions to form long oligomers.

The reaction was studied by taking samples of the mixture periodically during the polymerisation and monitoring the progress of the reaction through a variety of techniques including: SDS-PAGE, Analytical Size Exclusion Chromatography, Mass Spectrometry, Circular Dichroism and Transmission Electron Microscopy (TEM), these are discussed in the following sections.

3.3.8.1 Following the Polymerisation using SDS-PAGE

The most convenient way of monitoring the polymerisation reaction was using SDS-PAGE. This allows us to visualise the distribution of oligomers in the reaction mixture. To investigate the how the reaction proceeds a sample of His-Xa-CTPR3ΔS-thioester was initiated with Factor Xa and the reaction followed by removing samples at set times.

Figure 3.14 shows a reaction with time points taken at 0, 1 and 3 days. Prior to initiation, at 0 days, the sample contains the protein, His-Xa-CTPR3ΔS-thioester seen at 15 kDa. A small concentration of intein-CBD protein is visible at 27 kDa, this is a by-product of the purification of the His-Xa-CTPR3ΔS-thioester protein from the MESNa induced cleavage of His-Xa-CTPR3ΔS-intein-CBD. The majority of the intein-CBD protein remains bound to the chitin resin, however some is released along with the His-Xa-CTPR3ΔS-thioester. After addition of the Factor Xa

the His-tag is cleaved and this is shown by the disappearance of the band at 15 kDa. Also observed is the emergence of larger oligomers, seen at 17 kD (dimer) and 21 kDa (trimer). Further oligomers are seen at higher molecular weights. After 3 days of reaction the cleavage product has fully reacted (marked by the absence of the band previously seen at 12 kDa).

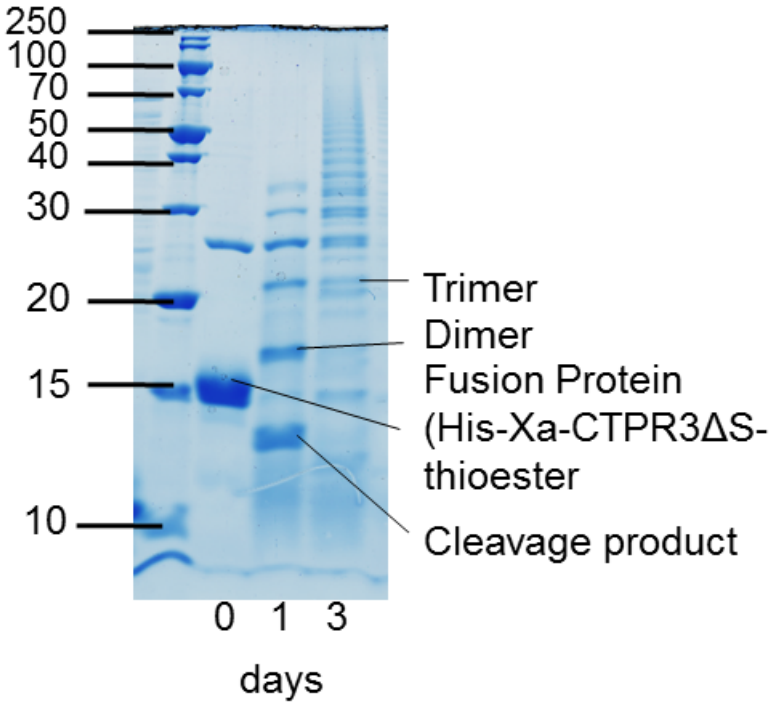

Figure 3.14: 18% SDS-PAGE gel of samples taken from a polymerisation reaction. The conditions were: 100 μM protein in 50 mM phosphate pH 7, 150 mM NaCl and 10 mM MESNa. Prior to initiation, at 0 days, the sample contains the protein His-Xa-CTPR3ΔS-thioester at approximately 15 kDa and a smaller concentration of the intein-CBD protein at 27 kDa. After addition of the Factor Xa the reaction proceeds for 1 day. At this point the His-tag has been fully cleaved to produce CTPR3ΔS-thioester, shown by the absence of a band at 15 kDa. Also observed is the emergence of larger oligomers, seen at 17 kD (dimer) and 21 kDa (trimer). Further oligomers are seen at higher molecular weights. After 3 days the cleavage product has fully reacted and the concentrations of the smaller oligomers has decreased as these units react to form larger structures.

To confirm the reproducibility and probe the finer details of the the polymerisation the reaction was repeated and samples removed at set time points. Samples were removed from the reaction at the following times: 0, 4, 8 and 12 hours and at 1, 2, and 7 days, Figure 3.15. The same trend is seen as before. As the Factor Xa cleaves

the His-tag, the band for the His-Xa-CTPR3ΔS-thioester (seen at 15 kDa) reduces in intensity. We can see the presence of a cleavage product that appears at 12 kDa, 4 hours after the initiation. After one day there is the emergence of dimers at 17 kDa and trimers at 25 kDa. After 2 days there are larger structures visible above 27 kDa, and the intensity of each band decreases as the size of the oligomerisation products become polydisperse.

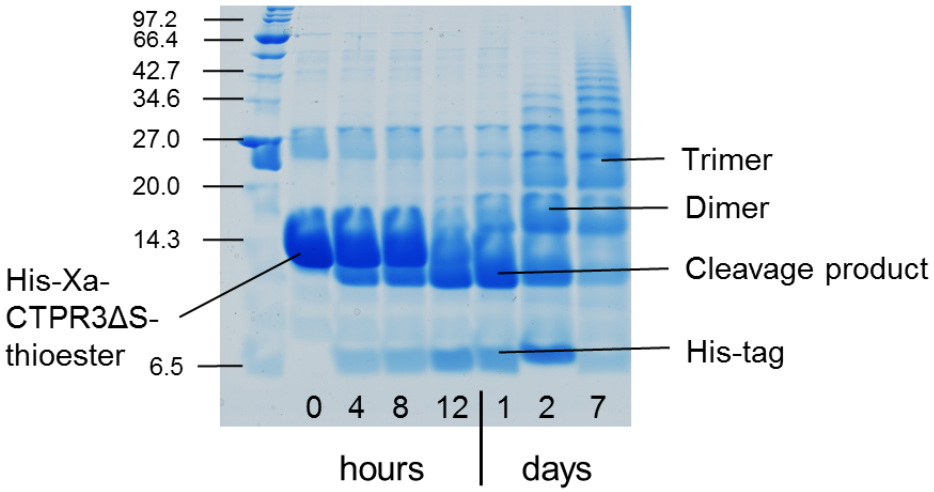

Figure 3.15: 18% SDS-PAGE gel of samples taken from a polymerisation reaction. The conditions were: 200 μM protein in 50 mM phosphate pH 7, 150 mM NaCl and 10 mM MESNa. The same volume of sample was loaded for each time point to allow a relative comparison of the concentration of each species in the sample. Prior to the addition of Factor Xa, at 0 hours, the His-Xa-CTPR3ΔS-thioester can be seen as a bright band at approximately 15 kDa. The intein-CBD protein is also present in a lower concentration at 27 kDa. As the Factor Xa cleaves the His-tag, the band for the His-Xa-CTPR3ΔS-thioester (seen at 15 kDa) reduces in intensity. We can see the presence a cleavage product that appears at 12 kDa, 4 hours after the initiation. In addition we see a band for the His-tag (seen at 6.5 kDa) increase in intensity as it is cleaved from the His-Xa-CTPR3ΔS-thioester protein. After one day there is the emergence of dimers at 17 kDa and trimers at 25 kDa. After 2 days there are larger structures visible above 27 kDa, and the intensity of each band decreases as the size of the oligomerisation products become polydisperse.

It is interesting to note that the bands on the gel appeared to be smaller than the predicted molecular weights. Table 3.1 shows the calculated molecular weights. Whilst there is a linear trend between the molecular mass and retention on the gel, there are several factors that can alter a proteins' mobility. One factor that

strongly influences how proteins migrate is how well they bind to SDS [107]. Proteins that bind poorly to SDS migrate faster through the gel matrix and appear to have a smaller molecular mass [107]. The CTPR proteins always appear at a smaller molecular mass on an SDS-PAGE gel than the calculated weight. Appendix Section 7.0.5 shows an example of how different length CTPRs appear on an SDS-PAGE gel. This trend has been observed consistently throughout the research conducted on CTPRs [Personal communication, Regan and Main research groups]. Previous studies have confirmed via mass spectrometry that various CTPRs are the intended mass, however they appear to be have a smaller mass as judged by SDS-PAGE. We conclude that the shift in the weights seen here for the oligomers is due to reduced binding of SDS. This gel-shift observation is consistent with that typically observed for CTPRs. Samples were analysed using MS and this confirmed that the proteins are the intended mass, see Section 3.3.8.3.

Table 3.1: Predicted molecular weights of polymerisation products

CTPR Species	**Molecular Mass (Da)[1]**
His-Xa-CTPR3ΔS-thioester-intein-CBD	44867
intein-CBD	27858
His-tag	4459
His-Xa-CTPR3ΔS-thioester	17126
CTPR3ΔS-thioester	12719
Dimer	25303
Trimer	37887

[1]Masses Calculated from the sum of the average isotopic mass of each amino acid.

3.3.8.2 Analytical Size Exclusion Chromatography

Size exclusion chromatography was used to analyse the size of species present in the polymerisation reaction. A polymerisation reaction was set up using 200 μM His-Xa-CTPR3ΔS-thioester in 50 mM phosphate pH 7 with 150 mM NaCl and 100 mM MESNa. Samples were removed and frozen at the following times: 0, 4 and 8 hours and at 1, 2, 5 and 6 days. The samples were defrosted, diluted down 1 in 4

and spun at 14 000 rpm for 10 minutes to remove any large aggregates. The samples were loaded onto a GE Healthcare Superdex 75pg 300/10 and the elution monitored using the absorbance at 280 nm..

Figure 3.16A shows the chromatograms of the time points: 4 hours (black), 2 days (blue) and 5 days (red). Inset are the chromatograms of CTPR3ΔS (green) and CTPRa6 (red). Figure 3.16B is a SDS-PAGE gel of the samples taken from the same polymerisation reaction as the samples shown in Figure 3.16A. The SDS-PAGE gel shows that this polymerisation did not proceed as successfully as previous reactions, such as that shown in Figure 3.14. There is still a high concentration of the CTPR3ΔS-thioester protein present in the later samples, showing that the protein hasn't oligomerised. We now know that the decreased reaction rate is likely to be because the His-Xa-CTPR3ΔS-thioester protein was stored before reaction rather than being used immediately after purification. Storage of the His-Xa-CTPR3ΔS-thioester protein was found to result in hydrolysis of the thioester leading to loss of reactively at the C-terminus, see Section 3.3.9.3 for further discussion. Despite this problem, the size exclusion chromatograms provide us with some insight into the polymerisation reaction, albeit at a slower reaction rate.

In the chromatogram in Figure 3.16A, there are two prominent peaks for the 4 hour time point: one at 10.0 ml and one at 9.2 ml with a lower intensity. We can identify these peaks by comparing their retention volumes with those of CTPR3ΔS (10.9 ml) and CTPRa6 (9.5 ml). These are more suitable size comparisons than standard size exclusion calibrants due to the elongated structures of CTPRs. CTPR3ΔS is approximately the size of the proteins His-Xa-CTPR3ΔS-thioester and CTPR3ΔS-thioester. Therefore the peak at 10.0 ml corresponds to the His-Xa-CTPR3ΔS-thioester and CTPR3ΔS-thioester proteins. CTPRa6, which elutes at 9.5 ml, is approximately the size of the dimer formed by the polymerisation reaction. Thus the peak at 9.2 ml corresponds to the dimer. Two days after initiation, the peak at

9.2 ml has increased in intensity signifying that there is an increase in the number of dimeric species. A shoulder also appears on the left side of this peak, which is the result of trimers being formed. After 5 days, the peak at 10.0 ml has shifted to 10.3 ml, this is the result of the cleavage of His-Xa-CTPR3ΔS-thioester into CTPR3ΔS-thioester. Also the intensity of this peak has reduced when compared with the earlier time point samples. The reduction in intensity is due to the decrease in concentration of the CTPR3ΔS-thioester protein as it oligomerises into larger species.

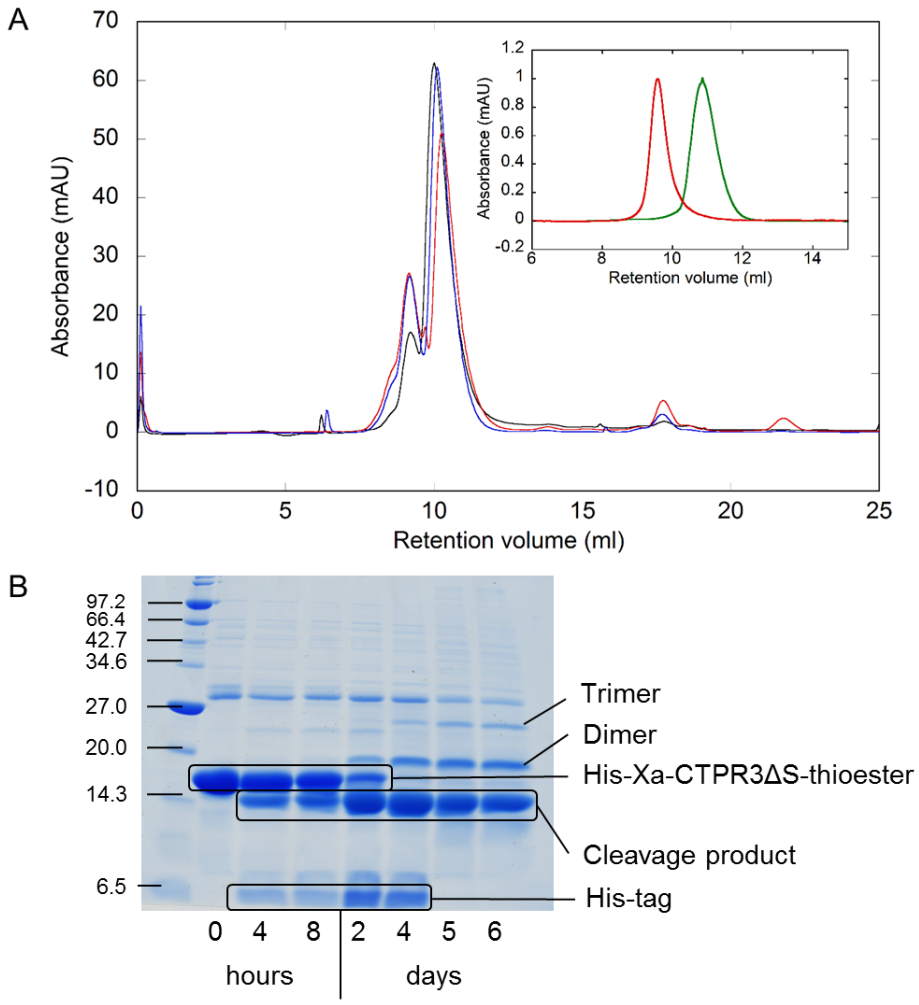

Figure 3.16: (A) Analytical size exclusion chromatograms for the time points: 4 hours (black), 2 days (blue) and 5 days (red). All 3 chromatograms show two prominent peaks: one at approximately 10 ml and a further peak with a lower intensity at 9.2 ml. The peak at 10 ml decreases in intensity as the reaction progresses. The peak at 9.5 ml increases in intensity between the 4 hour and 2 day time points. Inset are the chromatograms for CTPR3ΔS (green) and CTPRa6 (red) with retention volumes of 10.9 ml and 9.5 ml respectively. CTPR3ΔS is a similar size to the reactive monomer, CTPR3ΔS-thioester. CTPRa6 is close to the expected size (half a repeat larger) of the dimer. A GE Healthcare Superdex 75pg 300/10 column (equilibrated with 50 mM phosphate pH 7, 150 mM NaCl) was used. The flow rate was set at 0.5 ml/minute. The elution of the protein was monitored using the absorbance at 280 nm and the retention on the column measured in ml. (B) 18 % SDS-PAGE gel of time course samples used for the size exclusion chromatography. Prior to initiation, at 0 days, the sample contains the protein His-Xa-CTPR3ΔS-thioester at approximately 15 kDa and a smaller concentration of the intein-CBD protein at 27 kDa. As the Factor Xa cleaves the His-tag, the band for the His-Xa-CTPR3ΔS-thioester (seen at 15 kDa) reduces in intensity. We can see the presence a cleavage product that appears at 12 kDa, 4 hours after the initiation. In addition we see a band for the His-tag (seen at 6.5 kDa) increase in intensity as it is cleaved from the His-Xa-CTPR3ΔS-thioester protein. After 2 days there are dimers and trimers visible at 20 kDa and 26 kDa respectively.

Larger oligomers (> trimers) were not observed on the chromatograms, possibly because only a small concentration of these species were present in samples as judged by the SDS-PAGE gels. Another contributing factor is the inclusion of a centrifugation step in the preparation of the sample. This step was included in the protocol to protect the column from damage by particulates. However, high speed centrifugation will reduce the concentration of larger oligomers present in the sample and hinders a direct comparison with the SDS-PAGE gels.

3.3.8.3 Mass Spectrometry (MS)

MALDI-TOF mass spectrometry was used to confirm the mass of the His-Xa-CTPR3ΔS-thioester and to investigate the sizes of the products formed during the polymerisation reaction. All samples were prepared by desalting with a C18 zip-tip and were mixed in a 50 % ratio with sinapinic acid matrix. Figure 3.17 shows a sample of the His-Xa-CTPR3ΔS-thioester protein. The spectrum shows a peak at the expected mass, (17026 MW measured, 17129 MW calculated). Additional peaks at approximately 8.5 kDa and 6.4 kDa are the +2 and +3 ions respectively.

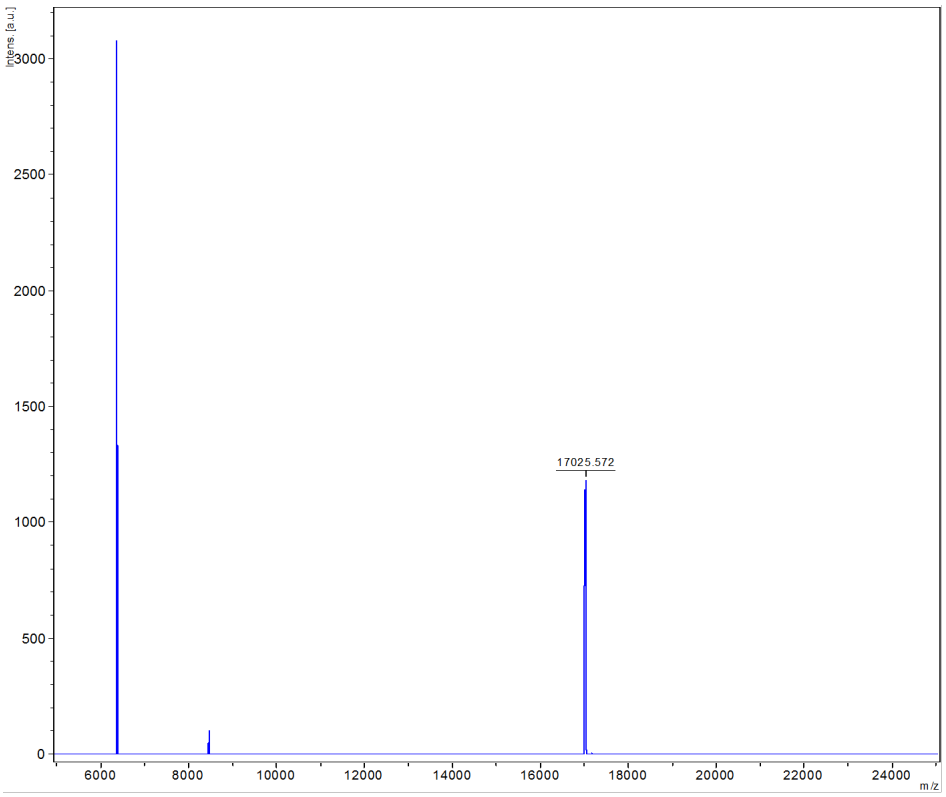

Figure 3.17: Mass spectrum of the His-Xa-CTPR3ΔS-thioester protein (17026 MW measured, 17126 MW calculated). Peaks seen at approximately 8.5 kDa and 6.4 kDa are the +2 and +3 ions respectively.

Figure 3.18 is a sample of a polymerisation reaction taken 6 hours after initiation. Here we can see the presence of the active monomer CTPR3ΔS-thioester (12873 MW measured, 12719 MW calculated). In addition there are dimers (25503 MW measured, 25303 MW calculated) and trimers (40859 MW measured, 37887 calculated) visible. We know from SDS-PAGE that there are oligomers larger than trimers produced by the polymerisation, however we were unable to observe these via MS. This is because the larger oligomers do not ionise from the matrix.

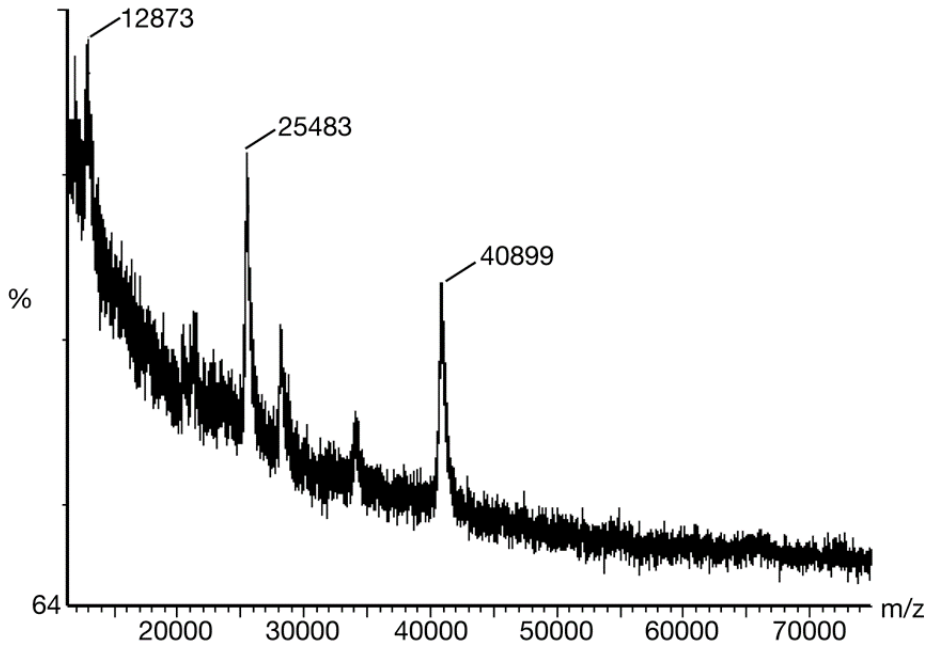

Figure 3.18: Mass spectrum of the polymerisation reaction after 6 hours. This spectrum shows the 3 significant species: active monomer CTPR3ΔS-thioester (12873 MW measured, 12719 MW calculated), dimer (25503 MW measured, 25303 MW calculated) and trimer (40859 MW measured, 37887 calculated).

Figure 3.19 is a sample of the reaction taken at 3 days, taken from the same polymerisation reaction as that shown in Figure 3.14. This spectrum showed peaks at 12 and 15 kDa (the smaller mass peaks are 2+ ions). The active monomer, CTPR3ΔS-thioester, is 12 kDa therefore we can conclude that Factor Xa cleavage is yielding the correct product. However, the presence of another peak at 15 kDa suggests that the Factor Xa is sometimes cleaving at 2 locations in the protein. A review of the literature revealed that Factor Xa can be promiscuous and frequently cleaves at alternative sites [108, 109].

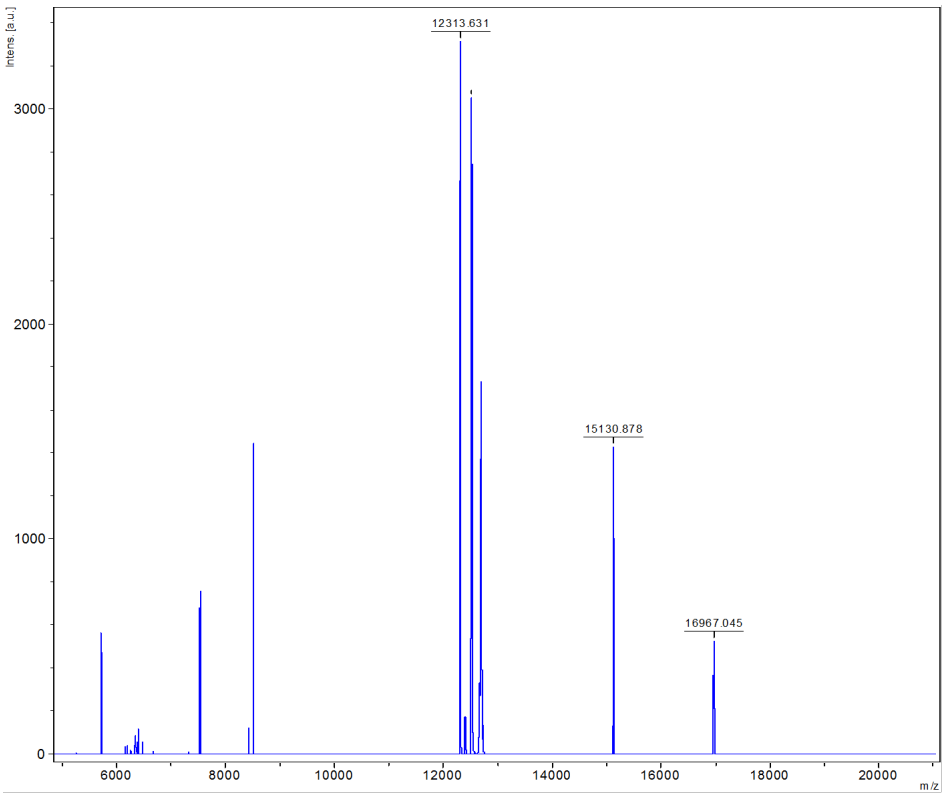

Figure 3.19: Mass spectrum of the polymerisation reaction after 3 days shows that the His-Xa-CTPR3ΔS-thioester protein (16967 MW measured; 17129 MW calculated) is cleaved by the Factor Xa at the desired cleavage site (12314 MW measured; 12719 MW calculated) and also at a second site to give a peak at (15131 MW). Peaks at 8.5 kDa and below are the +2 and +3 ions of the 3 species observed. It should be noted that the intensity of the peaks observed in MS are all relative and are not an indicator of the concentrations present in the sample. The mass spectra presented here is limited by the poor calibration of the machines, leading to variation between the calculated and the observed masses.

Figure 3.20 compares the designed site for Factor Xa cleavage (A) with a proposed alternative site (B). Only the intended cleavage site yields the N-terminal cysteine that is needed for NCL, therefore cleavage at the upstream site is unproductive and inhibits the oligomerisation.

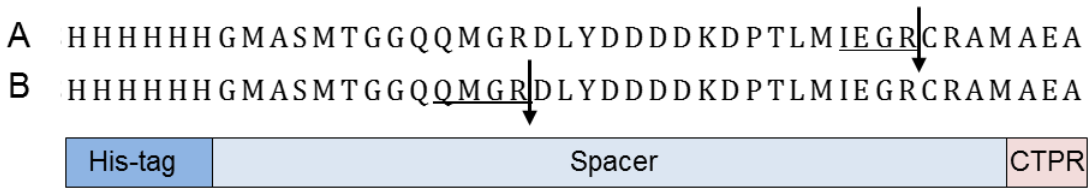

Figure 3.20: Comparison of intended Xa cleavage site (A) and proposed alternative site (B) at the N-terminal section of the full-length intein protein. The intended site shown in (A) leaves the N-terminal cysteine residue necessary for NCL. However, MS data suggests that there is cleavage at an upstream site as a larger species (15.1 kDa) was detected via this method. Reviews of the effectiveness of Factor Xa show that the enzyme is promiscuous and often cleaves at different sites from its normal site of IEGR↓. We propose that Factor Xa cleaves at the site shown in (B), as this would produce the same size fragment as was seen via MS.

These data confirms that the protein purified from the chitin resin is His-Xa-CTPR3ΔS-thioester. When we analysed samples of the polymerisation reaction we can see that cleavage of the His-tag results in oligomerisation to yield dimers and trimers.

3.3.8.4 Circular Dichroism

Circular Dichroism was used to assess the secondary structure of the monomer and the products of the polymerisation reaction. Circular Dichroism scans were measured for the same samples as those shown in Figure 3.16(B). Samples were diluted down by a factor of 16 (approximate protein concentration 12.5 μM) and the ellipticity of the samples was measured in a 1 mm cuvette.

The 0 day time point has a negative peak at 222 nm this confirms that the His-Xa-CTPR3ΔS protein is α-helical. There is an increase in the ellipticity at 222 nm that is observed as the reaction proceeds, with the 5 day sample displaying a more negative value. This is indicative of an increase in the helical content of the protein within the sample which is caused by the increased length of the CTPR oligomers. This confirms that the polymerised CTPR units are docking together and aligning

into an extended TPR fold, as designed.

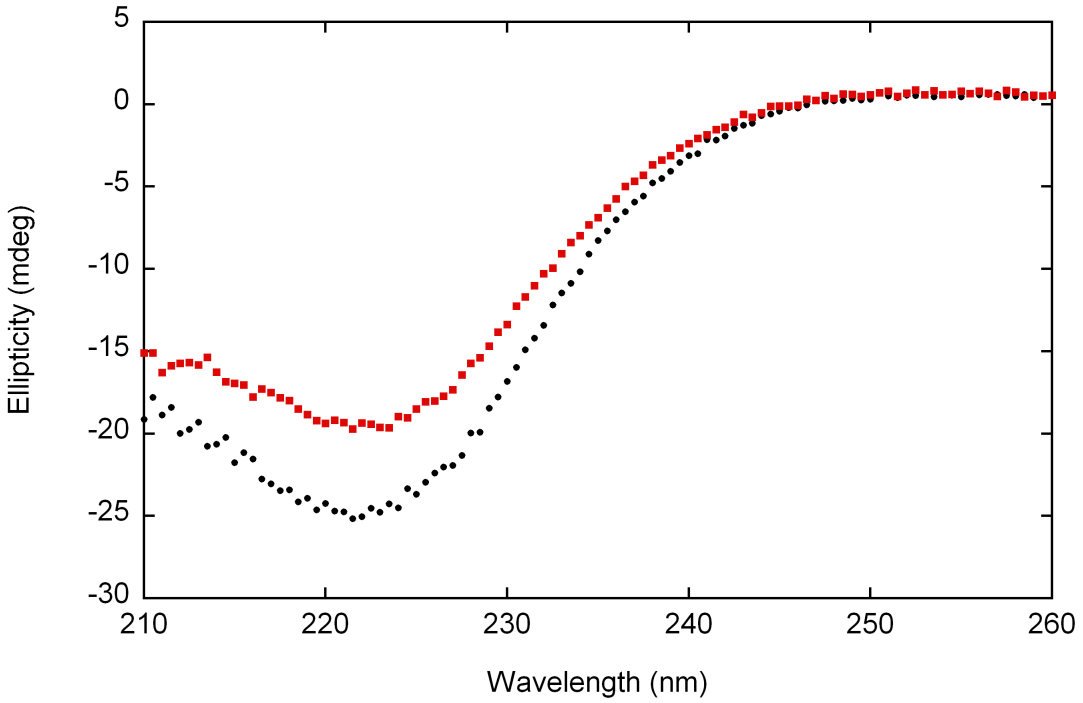

Figure 3.21: CD wavescans for the 0 day (red squares) and 5 day (black circles) time points. Conditions were: 50 mM phosphate pH 7, 150 mM NaCl, 10 mM MESNa, protein concentration was 12.5 μM and a 1 mm cuvette was used. There is an increase in the ellipticity at 222 nm as the reaction proceeds, indicating an increase in the helical content of the protein within the sample caused by the increased length of the CTPR oligomers.

3.3.8.5 Transmission Electron Microscopy (TEM)

Transmission electron microscopy is used to visualise biomolecules, typically the sample is applied to a carbon grid followed by a stain that coats the outside of the sample. We decided to use TEM to assess the morphology of the structures formed by the polymerisation reaction. A sample of His-Xa-CTPR3ΔS-thioester (100 μM in 50 mM phosphate pH 7, 150 mM NaCl, 10 mM MESNa) was initiated with Factor Xa and samples were removed at 2, 8, 12 and 96 hours. These samples were also analysed using SDS-PAGE, this confirmed that the polymerisation reaction proceeded in the same way as previous reactions. Samples were prepared for TEM by dilution to approximately 2 μM using ddH_2O. Samples were stained with

2% (w/v) uranyl acetate. Dilution was necessary because of the high protein concentration and the phosphate buffer. Phosphate buffers interact with uranyl stain to form crystalline precipitates that can obscure the sample. Glow Discharged Carbon Formvar 400 mesh grids were used.

The micrographs in Figures 3.22 and 3.23 are representative of the fibrous structures that were observed via TEM as a result of repeated NCL reactions. We observed these fibres on multiple grids prepared from the various time points. We also prepared samples from other polymerisation reactions to confirm the reproducibility of the fibre formation and the same fibrous structures were seen. In Figure 3.22 the progression of the NCL reaction can be observed. At 0 hours, 3.22A, there are no visible structures as the His-Xa-CTPR3ΔS-thioester protein is too small to be observed (>3 nm long) using TEM. However, after 2 hours, small granular species with a narrow size distribution are visible, Figure 3.22B. After 12 hours, Figure 3.22C, the micrographs show structures that were growing in size, gaining in aspect ratio such that some of them corresponded to short filaments. These species appear to be 7 nm across, but have differing lengths. By 96 hours, Figure 3.22D, the distribution of sizes had completely changed from the earlier time points. There was no clear modal size and many fibrous structures were observed. These include individual fibres that range in size up to 1 μm in length but are still only 7 nm across. Also visible are fibres that appear to have co-associated to form thicker filaments as judged by the increased width of the fibres. Inset in Figure 3.22D is an enlarged image showing the branch point between 2 fibres.

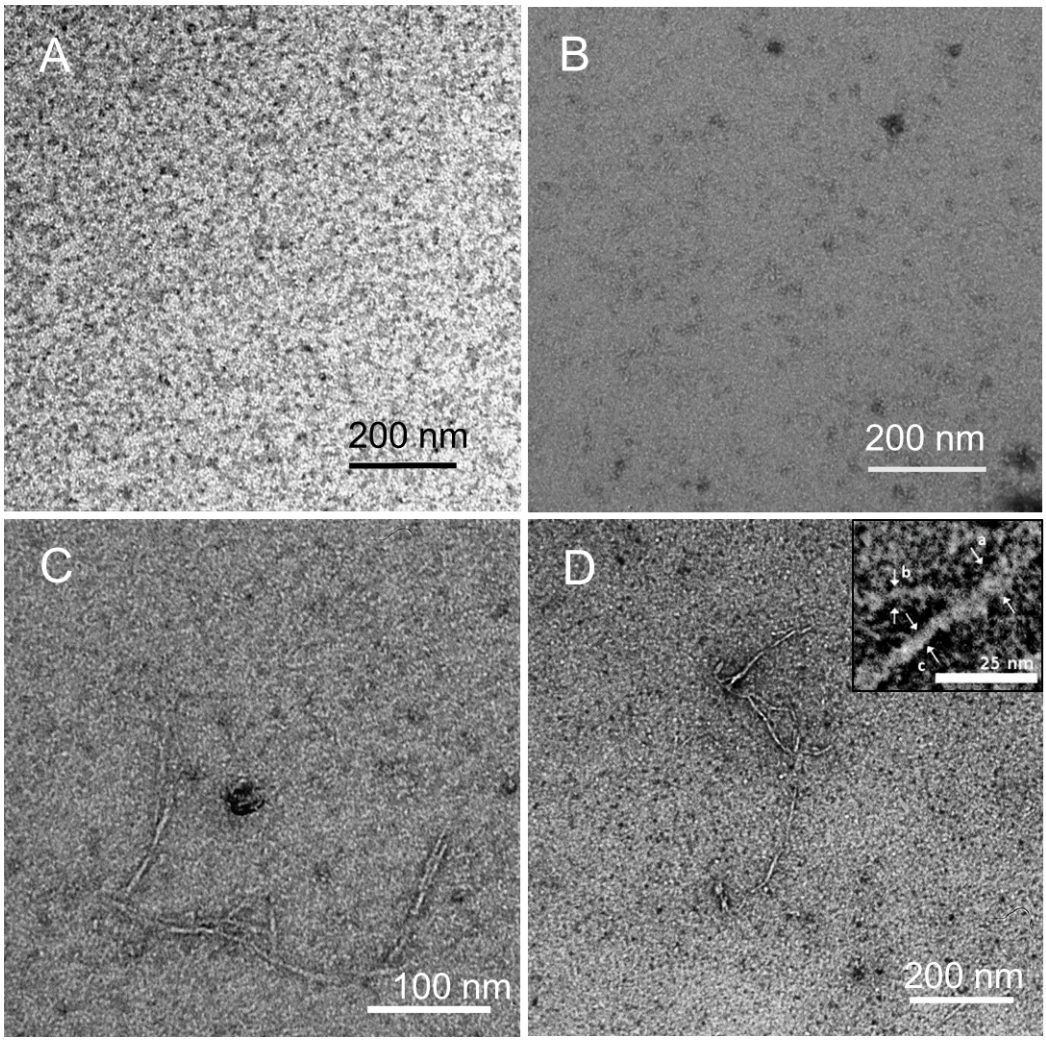

Figure 3.22: TEM images of samples of the polymerisation mixture (stained with uranyl acetate) removed from the reaction vessel at set time points: 0 hours (A), 2 hours (B), 12 hours (C), and 96 hours (D). (A) At the start of the reaction there are no visible structures as the monomeric units of His-Xa-CTPR3ΔS-thioester and the cleaved CTPR3ΔS-thioester are too small to be observed. After just 2 hours (B) there is the emergence of a narrow size range of small granular species. At 12 hours (C) there are fibres visible that are approximately 7 nm in width and variable lengths. After 96 hours (D) there are fibres up to microns in length. The width of the fibres is variable, some are 7 nm and some are wider, up to 17 nm wide. This could be the result of co-association of the fibres. Inset is a enlarged image of the 96 hour time point, showing a branch point between 2 fibres. Position (a) measures 17 nm where 2 fibres are associated, positions (b) and (c) measure 7 and 8 nm respectively. Samples were prepared for TEM by dilution to approximately 2 μM using ddH_2O. 2% (w/v) uranyl acetate stain was used.

The diameter of a CTPR superhelix was measured as 5 nm via X-ray crystallographic studies [110]. Our CTPR fibres are 7 nm in diameter when measured from the TEM images (the larger size is due to the negative stain). This close agreement shows

that the CTPR3ΔS-thioester protein oligomerises, as designed, to form fibres that are a single protein chain. We observed that these fibres are only linear for lengths of less than 30 nm, this suggests that the fibres can flexible or perhaps is due to improper docking of the CTPR3ΔS units following oligomerisation. Though flexible single filaments were predominantly observed, thicker fibres were also seen. Figure 3.23 is an example of the clustered filaments that were observed. They possessed diameters ranging up to 17 nm. This indicates that the individual filaments of CTPR polymer are not only flexible, but may also co-associate to form higher-order assemblies. However, the appearance of associated fibres could be an artefact of the staining process, the result multiple fibres are deposited on top of one another.

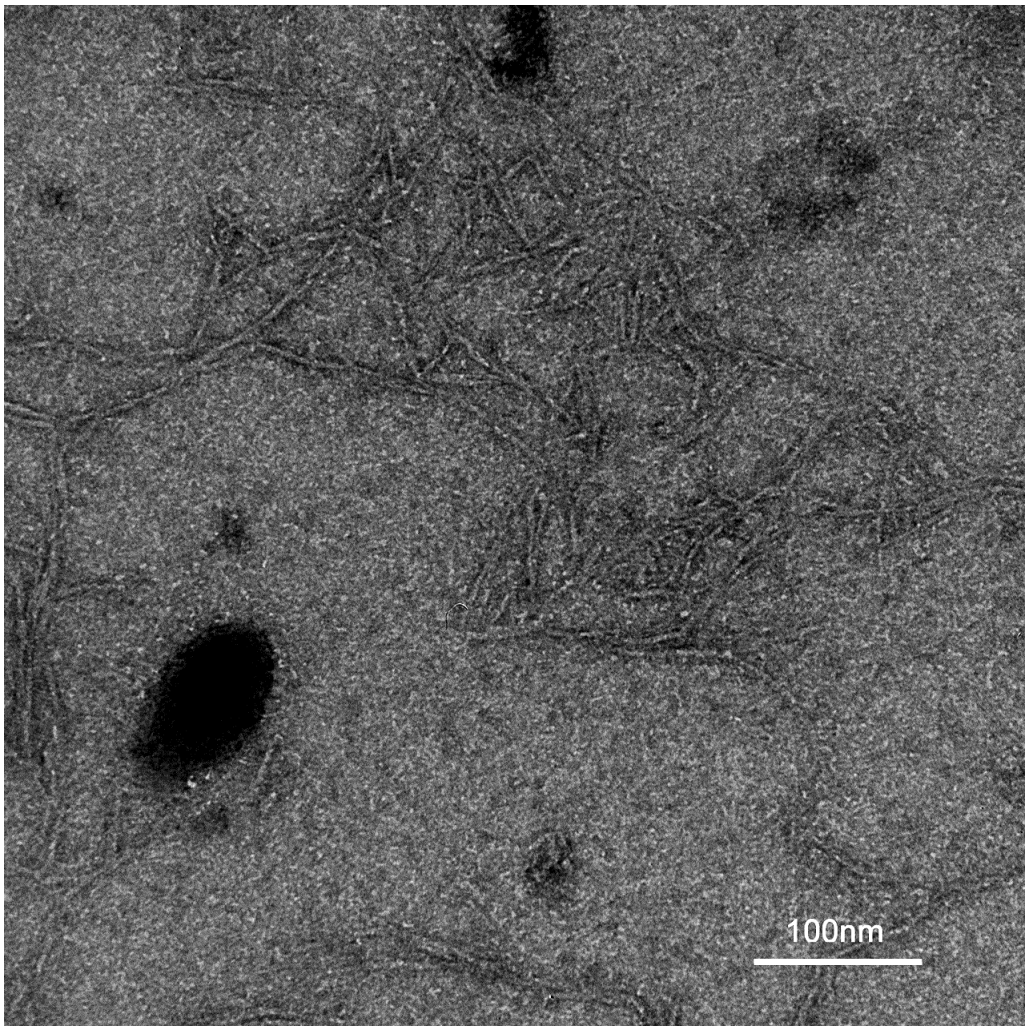

Figure 3.23: TEM image of a sample of polymerisation reaction taken after 8 hours after initiation. The width of an individual fibre is 7 nm, which is close to the width of a single CTPR superhelix (5 nm). Some fibres appear to co-associate to form thicker filaments, however this could be an artefact of the staining process. Samples were prepared for TEM by dilution to approximately 2 μM using ddH_2O. Samples were stained with 2% (w/v) uranyl acetate.

To ensure that any species observed on the TEM grids was the product of oligomerisation, grids were prepared using samples of the reaction buffer and none of these showed small or fibrous species. Also we observed that the micrographs at the start of the reaction do not show any fibrous structures. The fibres only appear after 8 hours confirming that they are the product of the polymerisation reaction. Although we did not perform this experiment a further control could use a mutant of the His-Xa-CTPR3ΔS-thioester protein lacking the crucial cysteine. This protein

could be cleaved by the Factor Xa but NCL would not occur and fibres should not be seen.

Using a representative sample of images from the time points: 2, 8, 12 and 96 hours, the distribution of sizes was calculated. The length of each fibre was recorded and expressed as a relative frequency of the total number of fibres that were imaged, Figure 3.24. From this we are able to see how the distribution of sizes changes as the reaction proceeds. At the start of the oligomerisation the range of sizes is narrow with only small species present. After 12 hours the average size increases and by 96 hours there is no clear average size.

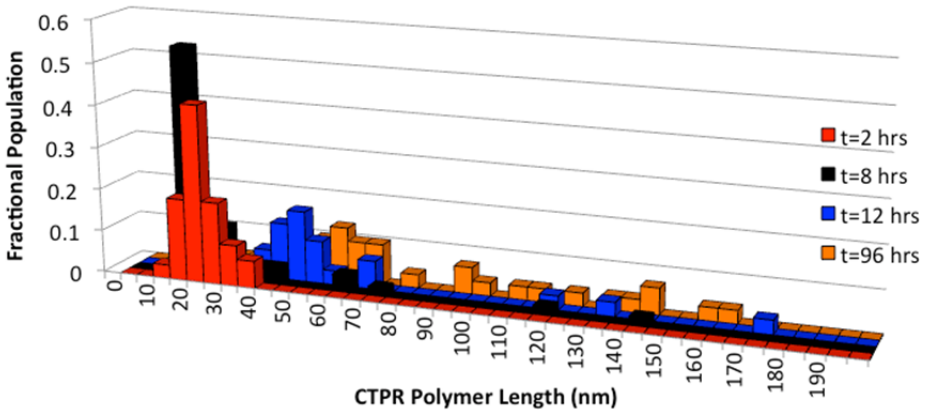

Figure 3.24: Relative frequency of CTPR oligomer lengths observed at time points 2, 8, 12 and 96 hours. The distribution of lengths of the growing CTPR oligomer was narrow after two hours, a similar distribution was observed after eight hours, with occasional detection of longer species. After 12 hours, a significant shift in the size of the oligomers was detected. After 96 hours, the distribution is completely different from the earlier the earlier time points with no clear modal size. Data were derived from measurements of TEM images, processed in EMAN2 and measured in GIMP (Graphical Image Manipulation Program). Analysis conducted and figure produced by Dr JJ Phillips.

3.3.9 Optimising the Fibre Growth Conditions

After assaying numerous polymerisation reactions we identified a number of factors that appeared to influence the polymerisation reaction. These include: (1) non-specific aggregation of the fibres/monomers; (2) rate of cleavage of the His-tag and (3) hydrolysis of the C-terminal thioester (leading to loss of reactivity). We decided to investigate how altering the buffer conditions affect factors (1) and (2). Factor (3) seemed to be largely be influenced by how quickly the His-Xa-CTPR3ΔS-thioester protein was used after purification from the chitin resin.

3.3.9.1 Aggregation

To test if aggregation was affecting the reaction, a sample was taken of a polymerisation reaction after 2 weeks. The sample was centrifuged at 14 000 rpm for 20 minutes and a visible pellet was observed. Samples of the supernatant and the pellet were prepared for analysis via SDS-PAGE. The gel in Figure 3.25 shows that there is a much higher percentage of larger species in the sample of the pellet verses the supernatant. This indicates that some of the larger species are aggregating and precipitating out of solution.

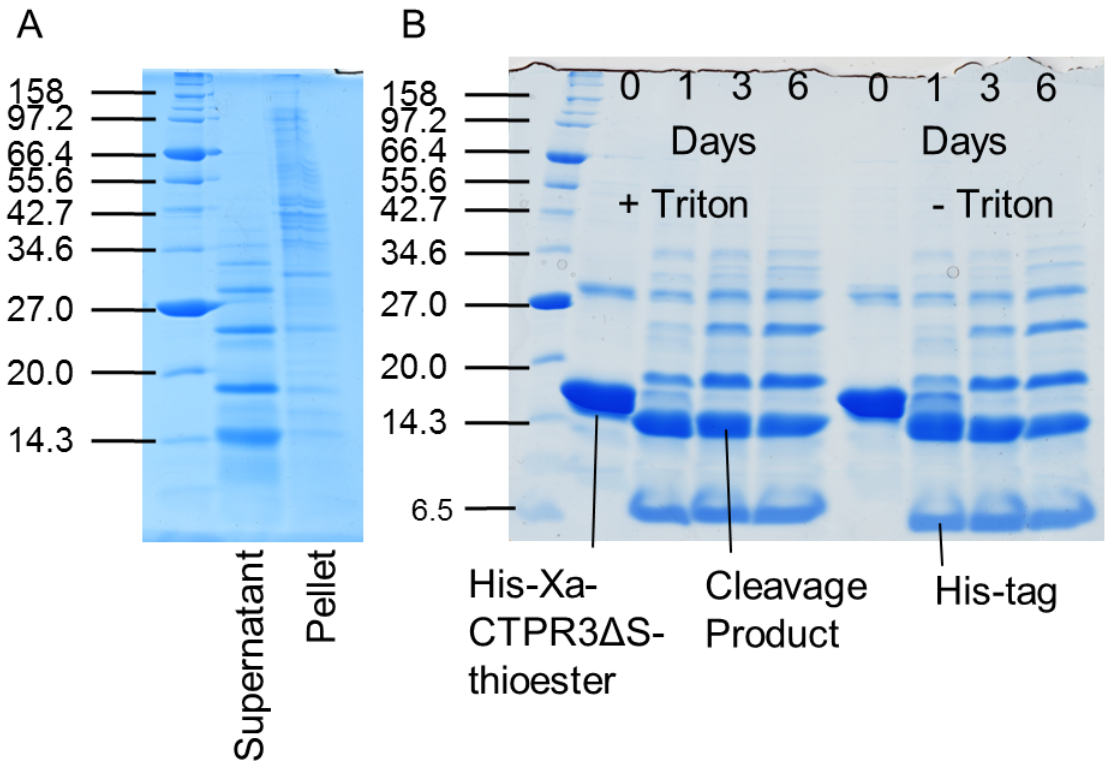

Figure 3.25: (A) 18% SDS-PAGE of the supernatant and the pellet from a sample of a reaction that had been progressing for 2 weeks. Conditions: 200 μM protein in 50 mM phosphate pH 7, 100 mM MESNa. (B) 18% SDS-PAGE gel of reaction with and without 1% triton in 50 mM phosphate pH 7, 500 mM MESNa. There is no discernible difference between the 2 samples.

Adding detergent can stabilise proteins and reduce aggregation, so a polymerisation reaction was set up with and without 1% triton in 50 mM phosphate pH 7, 500 mM MESNa. Several samples were taken over the course of 6 days and prepared for analysis via SDS-PAGE, see Figure 3.25. No difference was observed upon addition of 1% triton.

3.3.9.2 Effect of NaCl on the reaction

Observation of various polymerisation reactions seemed to indicate that addition of NaCl increased the speed of the reaction. To investigate this effect, an experiment was set up using 200 μM protein, in 50 mM phosphate 500 μM MESNa and either no

NaCl or 150 mM NaCl. After initiation, samples were removed for analysis via SDS-PAGE. In Figure 3.26B it is possible to see that after a month there is a significant increase in the number of higher order species in the 150 mM NaCl reaction. This shows that the reaction is more successful in the presence of NaCl. The increase in reaction rate may be due to increased reactivity of the Factor Xa, as the suggested buffer for the enzyme includes 150 mM NaCl.

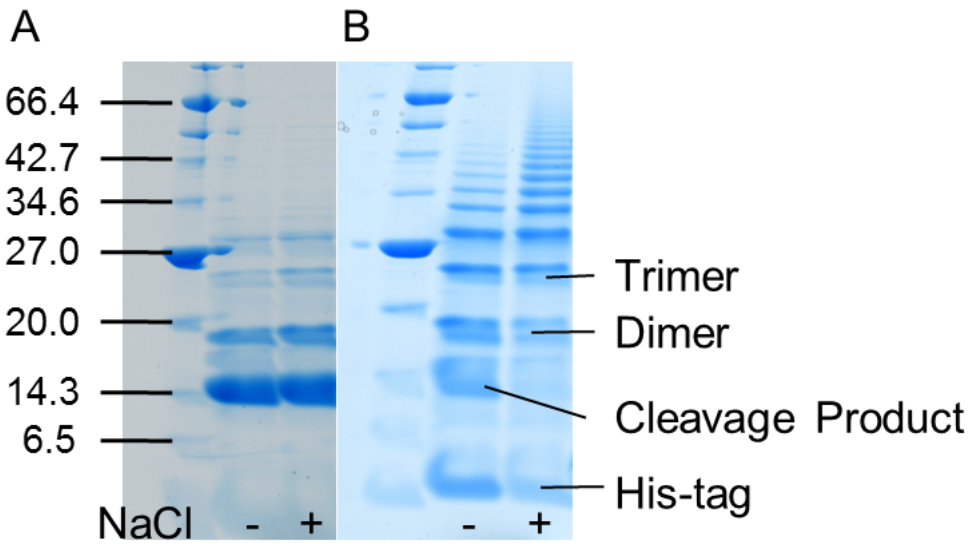

Figure 3.26: 18% SDS-PAGE gel of the polymerisation reactions performed with and without 150 mM NaCl after 1 week (A) and after 1 month (B). After 1 week there appears to be no difference between the samples. However after 1 month there is a significant increase in the number of higher order species in the 150 mM NaCl sample.

3.3.9.3 Hydrolysis of the thioester

The thioester at the C-terminus is susceptible to hydrolysis and this process is irreversible. Therefore once hydrolysis has occurred NCL cannot take place, resulting in the deactivation of that C-terminus. In some reactions the hydrolysis of the thioester resulted in a significant proportion of the CTPR3ΔS-thioester protein remaining unreacted, Figure 3.16B. Although buffer conditions do affect the rate of formation of fibres, the most successful reactions involved initiating the polymerisation as soon as possible after cleaving the His-Xa-CTPR3ΔS-thioester protein from the chitin resin.

Also important is adding enough Factor Xa to cleave the His-tag within 24 hours. For example the polymerisation reaction shown in Figure 3.14, involves a 'fresh' sample of His-Xa-CTPR3ΔS-thioester. Here the reaction has produced many larger structures after 3 days. However, the reactions shown in Figure 3.25B and Figure 3.16B used His-Xa-CTPR3ΔS-thioester protein that had been stored at 4 °C for 2 days and 5 days respectively after elution from the chitin column. Both of these reactions show a much slower rate with a higher concentration of the cleavage product present at 12 kDa and fewer larger oligomers.

3.4 Conclusions

Our aim was to self-assemble CTPR3ΔS proteins into larger structures and we have shown that this was successful. We have shown that our CTPR3ΔS-thioester monomers successfully and reproducibly undergo multiple rounds of NCL to yield oligomeric species, some of which are long enough to form fibres. Using TEM, we were able to visualise the fibres that were 7 nm wide and up to microns in length. In addition, the reaction only occurs when triggered by the cleavage of the His-tag by Factor Xa. The fibres can be formed under mild conditions: near neutral pH, at room temperature and in a variety of buffers. NCL has been used to synthesise oligomeric species such as collagen from short peptide chains [111]. Our system is the first to employ NCL in the formation of high order oligomers using recombinant repeat protein monomers.

The oligomerisation reaction was followed using TEM, SDS-PAGE, size exclusion chromatography and MS. The combination of these techniques gives us an understanding of the heterogeneous mix of products that are formed. TEM cannot show the monomeric species as it is below the detectable size (< 3 nm) but was used to show that fibrous structures were formed following repeated NCL. When stained with coomassie, SDS-PAGE can detect protein amounts as low as 100 ng allowing us to see very small concentrations of the higher order oligomers (>tetramers). Our size exclusion protocol included centrifugation of the samples prior to injection to avoid damaging the column by loading particulates. However this will alter the distribution of sizes in the sample as the very large oligomers will be removed for the sample, see Section 3.3.8.2 for more discussion.

When observed via TEM, the CTPR fibres were 7 nm in diameter, which is in close agreement with the diameter of a single CTPR superhelix seen by X-ray crys-

tallographic studies [8, 105] (once negative stain is accounted for). This shows that the CTPR3ΔS-thioester units oligomerise, as designed, via defined head-to-tail association to form single protein chain filaments. Interestingly, these filaments are only measurably linear for distances of less than 30 nm. This is in contrast to SAXS measurements that show CTPR proteins of up to 20 consensus TPR motifs are very rigid [55]. Flexibility beyond this length suggests that the CTPR polymer has some degree of freedom. This may be attributed either to a lack of rigidity in the superhelical structure or to incomplete docking together of units following covalent linkage.

To conclude, we have presented the synthesis by Native Chemical Ligation of a novel fibrous biopolymer from designed consensus repeat protein monomers. These structures have the potential to be used in the construction of biomaterials as they are both (1) formed from simple, soluble modular building blocks recombinantly expressed in large quantities and (2) each monomer contains a putative pentapeptide binding to allow functionalisation. The CTPR polymers thus represent a system that could be adapted for synthesis of nano- or micro-scale assemblies, decorated through binding of peptides with high affinity. One immediate application would stem from the recent interest in the use of TPR proteins to form hydrogels by constructing 18 CTPR motif sequences by recombinant approaches [55]. Our system allows the synthesis of much larger structures than those that can be formed recombinantly, offering the possibility of forming hyrdogels that have different properties. With a naturally occurring library of TPR protein scaffolds, each with their specific cognate peptide ligand, there are many opportunities for exploiting these self-assembling fibres by decorating them with peptide linkers.

Chapter 4

Introducing a Thermodynamic Switch into CTPR2 and CTPR3

In this chapter I will describe the characterisation of a series of CTPR deletion mutants and the introduction of a thermodynamic switch into CTPR2 and CTPR3.

4.1 Experimental Approach

A range of deletion mutants were produced and their stability was assessed using chemical equilibrium unfolding data. This data was fit using a Heteropolymer Ising model to analyse the multistate characteristics of the equilibrium unfolding. This model was then used to simulate the effect of introducing a thermodynamic switch into CTPR3, to selectively unfold the C-terminal helix, termed 'S-Switch'. This mutant (CTPRy3) was cloned, expressed and purified. Its ability to undergo a partial unfolding event was assessed using chemical and thermal denaturation data and 2D NMR. The equivalent mutation was introduced into CTPR2 and this mutant (CTPRy2) was assessed and compared with CTPRy3.

4.2 Motivation for Designing a Thermodynamic Switch

A **T**etratrico**P**eptide **R**epeat or TPR contains 34 amino acids that form a helix-turn-helix motif; each alpha helix is approximately 14 amino acids long [8], Figure 4.1. Several repeats stack together to form a tandem array and the hydrophobic core is capped at the C-terminus by a solvating helix (S helix). In Chapter 3 we have shown how CTPRs have been used to create fibrous nanostructures, where CTPR3ΔS monomers were triggered to react together in a specific head-to-tail manner. This method utilises enzymatic cleavage to induce the polymerisation.

A different approach to polymerising CTPR units could involve a specific unfolding event triggered by environmental factors that results in a compatible dimerisation interface being exposed. This could be made possible as the CTPR proteins have been shown to exhibit multistate folding [9, 78, 112], and therefore have access to partially folded stable intermediates. Although only small populations of these intermediates are ever present at a time, through engineering this percentage could be increased. The terminal helices are the best target for redesign into a switchable unit as this requires the least amount of structural change leaving the majority of the protein folded. To achieve this a series of deletion mutants were designed and their stability assessed using a Heteropolymer Ising model. This allowed us to simulate the population of partially unfolded intermediates at a given concentration of denaturant. This led to the design of the mutants CTPRy3 and CTPRy2, which undergo a partial unfolding event to reveal a polymerisation interface. These constructs unfold their C-terminal α-helix at a specific concentration of GuHCl. These intermediates are equivalent to a CTPRnΔS protein and possesses oligomerisation interfaces at both termini.

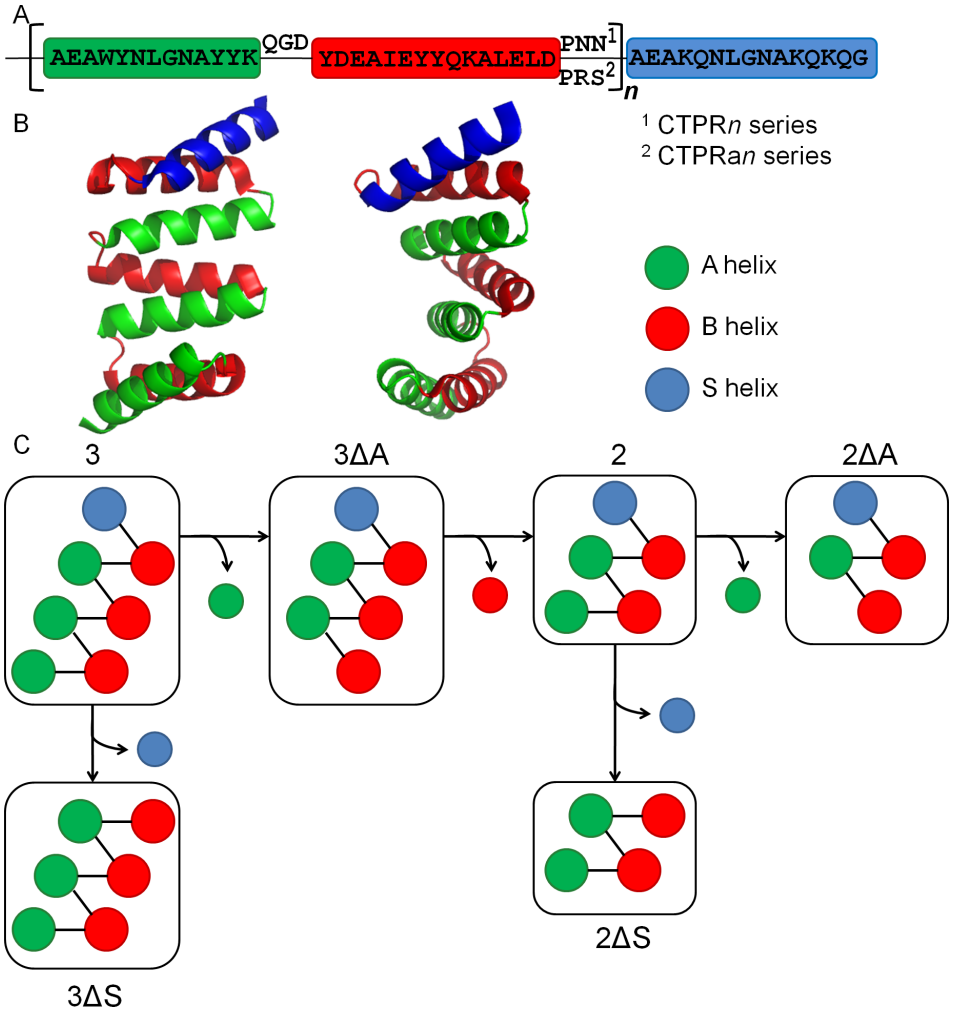

Figure 4.1: (A) A CTPR repeat is made up of two α helices (A and B), separated by a short loop, this sequence is repeated n number of times before there is a C-terminal capping helix, this is referred to as the 'solvating helix' or 'S helix'. There are 2 CTPR series: CTPRn and CTPRan. The sequences are the same except the last 2 residues in each repeat (positions 33 and 34) are NN in the CTPRn series and RS in the CTPRan series. (B) Crystal structure of CTPR3 (1NA0), with the helices coloured according to their sequence (A=green; B=red and C=blue). (C) Schematic showing the relationship and nomenclature between the mutants. Mutating out the N-terminal A helix produces ΔA. Mutating out the C-terminal helix produces ΔS.

4.3 Results

4.3.1 Design of Deletion mutants

Previous studies have shown that the stability of a CTPR increases with repeat number [9]. However, the effect of removing half of a repeat, i.e. 1 helix, had not been investigated. Thus to assess this, a series of mutants was devised, Figure 4.1 C. By studying the effect of removing terminal helices we will be able to determine the contribution to stability and thus determine the most suitable termini to situate a thermodynamic switch. The ΔA mutants lack the first 16 residues, that form the N-terminal helix. The ΔS proteins are truncated at position 34 and lack the C-terminal helix.

There are two series of CTPR proteins, CTPRn and CTPRan, n is the number of repeats. They possess almost identical sequences, except in each repeat, the last 2 residues (positions 33 and 34) are NN in the CTPRn series and RS in the CTPRan series 4.1 A. The CTPRan series of proteins is less stable when compared to CTPRn proteins of the same length as changing NN to RS at positions 33 and 34 results in the loss of a stabilising interaction between asparagine and a tyrosine in the A-helix of the preceding repeat, Figure 4.2.

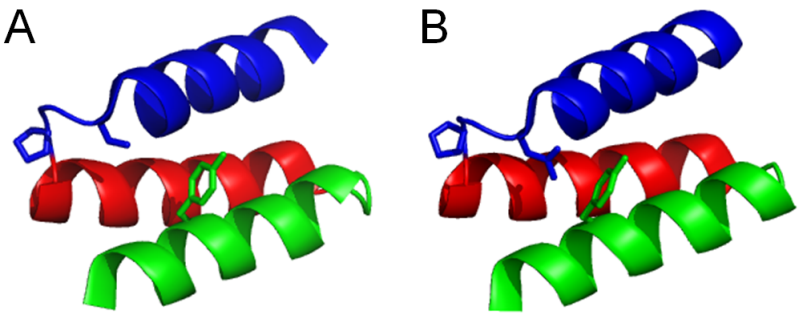

Figure 4.2: (A) Crystal structure showing the last 3 helices of CTPRa*n* (2F07). The final loop before the C-terminal S helix has the sequence PRS. The helices are coloured according to the sequence: A helix (green), B helix (red) and S helix (blue). (B) Crystal structure showing the last 3 helices of CTPR*n* (1NA0). The final loop before the C-terminal S helix has the sequence PNN. The interaction between the asparagine and the π system of the tyrosine residue stabilises the CTPR*n* series relative to the CTPRa*n* series.

4.3.2 Cloning of Deletion Mutants

Deletion mutants were cloned as described in Section 2.1 and Table 2.1. ΔA proteins were created using the PCR amplification technique described in section 2.2.1. ΔS proteins lack the C-terminal helix and were created using Site-directed Mutagenesis as described in Section 2.2.2. The mutations were confirmed via DNA sequencing.

4.3.3 Expression and Purification of Deletion Mutants

The proteins were expressed and purified under native conditions as described in Chapter 2.3.2. All give high yields of approximately 10 mg/L of culture.

4.3.4 Characterisation of Deletion Mutants

All proteins were monomeric as judged by the chromatograms for the size exclusion purification (data not shown). Figure 4.3A shows the far UV Circular Dichro-

ism wavelength scans of the deletion mutants displayed a negative peak at 222 nm, which indicates that they are α helical, as is expected for a CTPR. The molar ellipticity (expressed as mdeg $cm^{-2}dmol^{-1}$) for each protein was as follows: CTPR2 -1.72x10^5; CTPR2ΔS -1.31x10^5; CTPR2ΔA -1.37x10^5; CTPR3 -2.44x10^5; CTPR3ΔS -2.06x10^5 and CTPR3ΔA -2.00x10^5. The data shows that the ellipticity of those mutants that consist of the same number of helices, for example CTPR2ΔS and CTPR2ΔA, exhibit the same ellipticity. Also the ellipticity of the deletion mutants is lower than the full length an amount that equates to 1 helix.

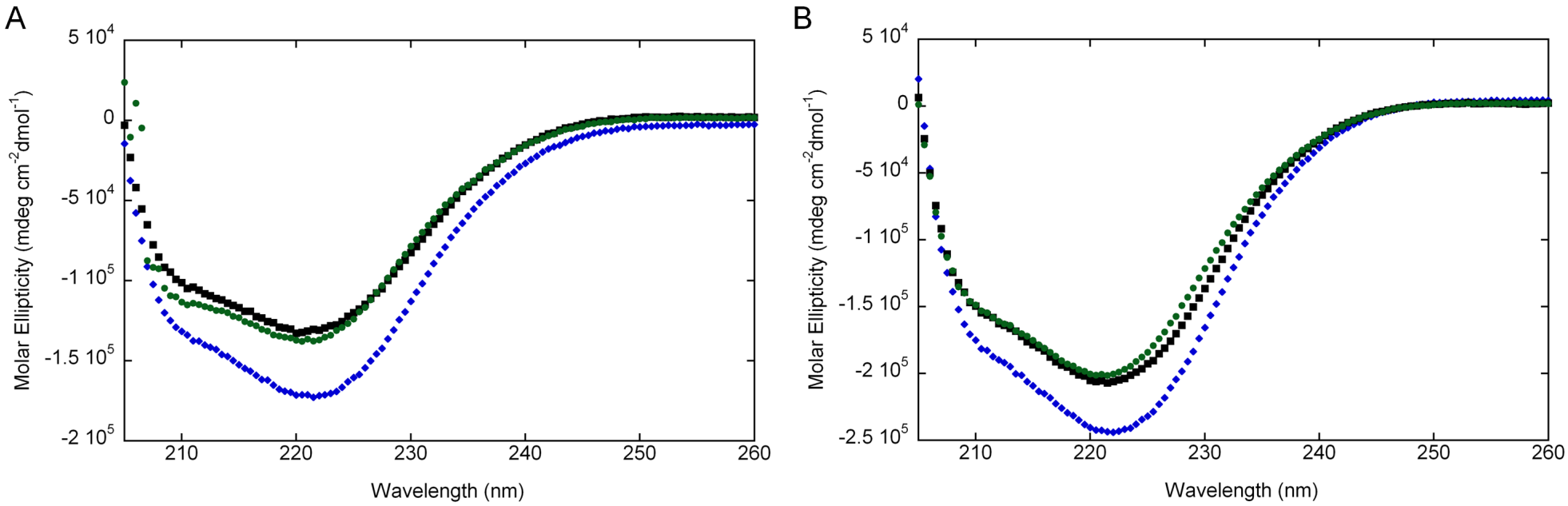

Figure 4.3: (A) Far UV wavelength scans of CTPR2 (blue), CTPR2ΔA (green) and CTPR2ΔS (black). (B) Far UV wavelength scans of CTPR3 (blue), CTPR3ΔA (green) and CTPR3ΔS. The samples were in 50 mM phosphate pH 7 and were held in a thermostated cuvette holder at 10 °C. The scan showed a single negative peak at 222 nm, indicating that the protein is highly α-helical.

4.3.5 Equilibrium Unfolding Studies

The series of deletion mutants were unfolded using GuHCl, the change in secondary structure was monitored through the change in CD signal at 222 nm. These experiments were conducted to assess the stabilities of the deletion mutants and compare the effect of removing the different terminal helices.

4.3.5.1 Chemical Denaturation

GuHCl denaturations were performed as described in Section 2.5.2.2. All the proteins underwent a single transition from the folded to unfolded state, Figures 4.4 and 4.5. The transition occurred at higher GuHCl concentrations as the number of helices was increased. The CTPRan series exhibited transitions at lower GuHCl concentrations than the equivalent size mutants form the CTPRn series. CTPR proteins have been shown to undergo reversible unfolding [9] and therefore refolding experiments were not conducted. The denaturations were performed in 50 mM phosphate pH 7. The samples were measured in a 5 mm pathlength cuvette held at 10 °C in a thermostated cuvette holder.

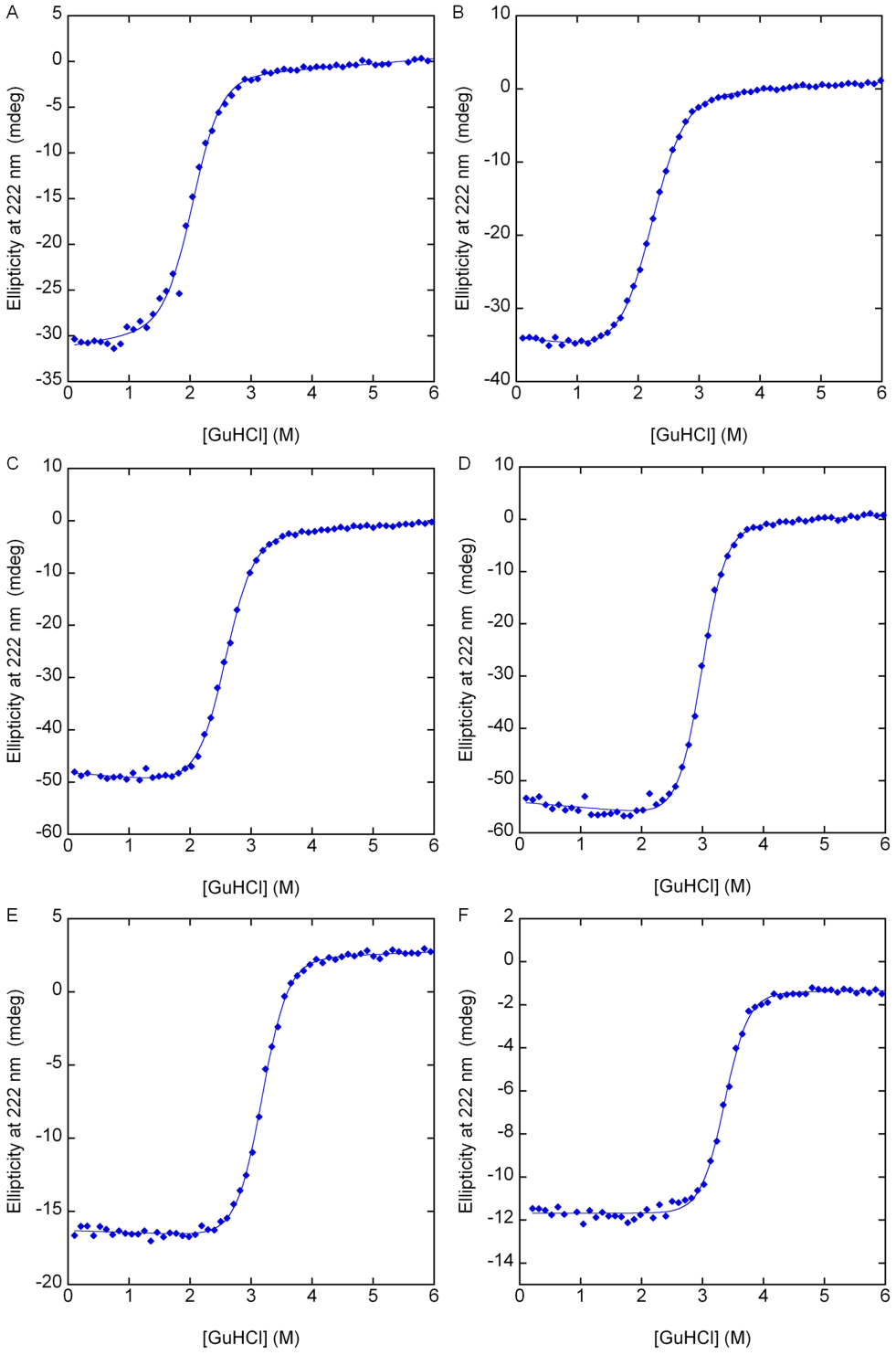

Figure 4.4: Chemical denaturation curves for CTPR2ΔA (A), CTPR2ΔS (B), CTPR2 (C), CTPR3ΔA (D), CTPR3ΔS (E) and CTPR3 (F).

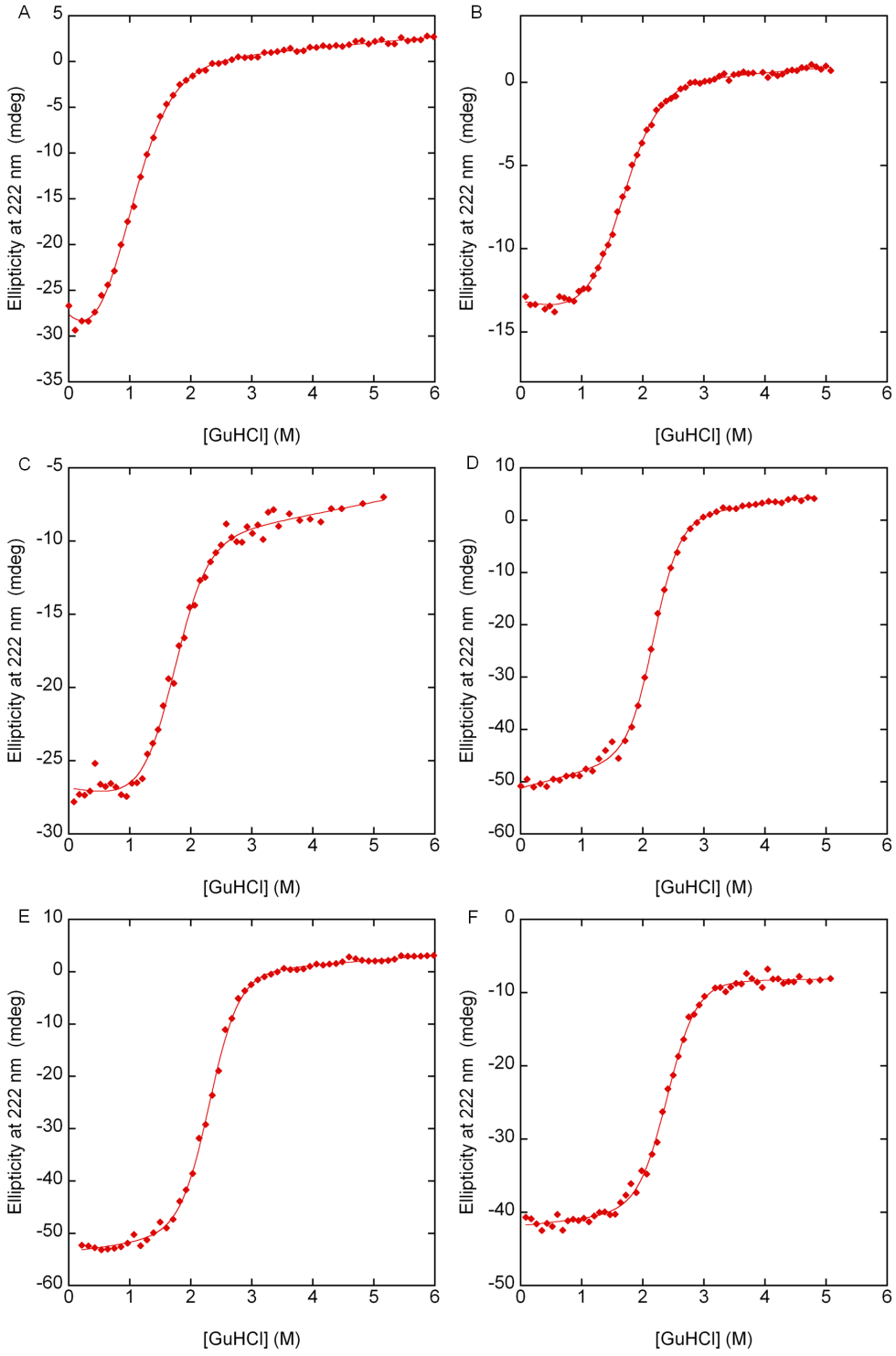

Figure 4.5: Chemical denaturation curves for CTPRa2ΔA (A), CTPRa2ΔS (B), CTPRa2[1] (C), CTPRa3ΔA (D), CTPRa3ΔS (E) and CTPRa3[1] (F). [1]Data collected by Dr Javadi

4.3.5.1.1 Normalisation of Chemical Denaturation Data

The CD signal increases with repeat number making it hard to visually compare equilibrium curves of mutants that are different lengths. Therefore to allow easier comparison of data sets, the equilibrium curves were normalised using Equation 4.1. Figure 4.6, shows the normalised equilibrium curves.

$$\lambda_{norm} = \frac{\lambda_{222} - \alpha_N}{\alpha_D - \alpha_N} \tag{4.1}$$

The values for α_D and α_N are the y-intercept values of the unfolded and folded baselines. This equation allows for the data to retain the slope of the folded and unfolded baselines.

4.3.5.1.2 Two-state Model Analysis of Chemical Denaturation Data

The simplest model to explain the thermodynamics of the unfolding of a protein is the two-state model. This states that the protein undergoes a single reversible transition from the folded to the unfolded state ($N \leftrightharpoons D$). Proteins can be qualitatively compared through comparison of the mid-point of denaturation, this value can be obtained through fitting the data to a two-state model, Equation 4.2. Although the unfolding of CTPR proteins is better explained through multistate unfolding [9], as the data shows a single transition the data can be fit to a two-state model.

$$\lambda_{obs} = \frac{(\alpha_N + \beta_N[D]) + (\alpha_D + \beta_D[D])\, exp^{\left(-\frac{m([D]-[D]_{50\%})}{RT}\right)}}{1 + exp^{\left(-\frac{m([D]-[D]_{50\%})}{RT}\right)}} \tag{4.2}$$

λ_{obs} is the ellipticity at 222 nm and $[D]$ is the concentration of GuHCl at any given λ_{obs}. Here α_N and α_D are the y-intercepts and β_N and β_D are the gradients of the folded and unfolded baselines respectively. The gradient through the transition between folded and unfolded is defined as m. The m value is a constant of pro-

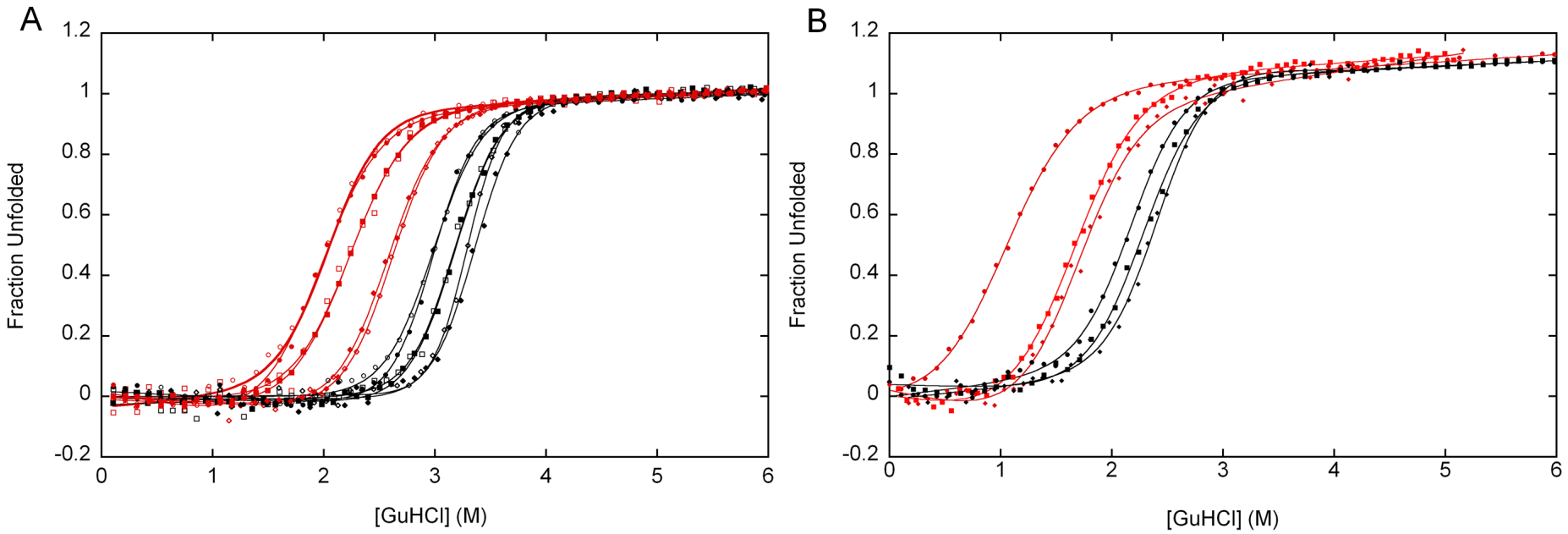

Figure 4.6: A) Normalised chemical denaturation curves for the CTPRn series CTPR2 (red diamonds), CTPR2ΔS (red squares), CTPR2ΔA (red circles), CTPR3 (black diamonds), CTPR3ΔS (black squares) and CTPR3ΔA (black circles). (B) Normalised chemical denaturation curves for the CTPRn series CTPRa2 (red diamonds), CTPRa2ΔS (red squares), CTPRa2ΔA (red circles), CTPRa3 (black diamonds), CTPRa3ΔS (black squares) and CTPRa3ΔA (black circles). All proteins undergo a single reversible transition from the folded to unfolded state. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

portionality related to the change in solvent accessible surface area (SASA) of the protein as it transitions between the native and unfolded state. At the midpoint of denaturation ($[D]_{50\%}$) in a two-state approximation, half of the population of the protein molecules are folded and half are unfolded. To calculate $[D]_{50\%}$ and m, the spectroscopic signal (λ_{obs}) was plotted against the denaturant concentration in Kaleidagraph version 4 (Synergy Software). The data was fitted to equation 4.2 using linear regression analysis in Kaleidagraph. Once these parameters are known, they can be used in Equation 4.3 to calculate $\Delta G^{H_2O}_{D-N}$, which is the change in free energy of unfolding in water. Data shown in Table 4.1.

$$\Delta G^{H_2O}_{D-N} = m[D]_{50\%} \tag{4.3}$$

4.3.5.1.3 Comparison of the Mid-points

The mid-point of denaturation was calculated from the fit of the equilibrium unfolding curve for each experiment, Table 4.1. The mid-point of denaturation can be used as a way of comparing the stability of different proteins. The data showed that the stability increased with repeat number. However, mutants with the same number of helices didn't have the same stability. This shows that all helix deletions were destabilising relative to their wild type proteins, with the greatest shift in midpoints and thus destabilisation occurring on removal of the N-terminal A helix rather than the C-terminal S helix. Therefore the N-terminal helix must contribute more favourable interactions than the C-terminal helix. This is most likely due to a combination of the difference in the primary sequence and the packing of the helices. Both the N-terminal and C-terminal helices are based on the same consensus sequence, however, the C-terminal helix was modified in the original design to increase solubility. This was achieved by replacing all the solvent facing hydrophobic residues to hydrophilic counterparts: `AEAWYNLGNAYYKQG` vs. `AEAKQNLGNAKQKQG`, respectively [8]. In addition both have different packing interactions which cause the C-terminal S helix to be 25

Table 4.1: Thermodynamic parameters for the GuHCl induced unfolding of CTPR proteins in phosphate buffer

Construct	**Data Set 1**			**Data Set 2**		
	[a] $[D]_{50\%}$ (M)	[a] m-value $kcalmol^{-1}M^{-1}$	[b] $\Delta G^{H_2O}_{D-N}$ $kcalmol^{-1}$	[a] $[D]_{50\%}$ (M)	[a] m-value $kcalmol^{-1}M^{-1}$	[b] $\Delta G^{H_2O}_{D-N}$ $kcalmol^{-1}$
CTPR2ΔA	1.94±0.01	2.08±0.05	4.0±0.1	1.99±0.02	2.08±0.08	4.1±0.2
CTPR2ΔS	2.21±0.01	2.10±0.03	4.6±0.1	2.24±0.02	2.18±0.15	4.9±0.3
CTPR2	2.57±0.01	2.43±0.06	6.2±0.2	2.62±0.01	2.55±0.06	6.7±0.2
CTPR3ΔA	2.97±0.01	2.82±0.10	8.4±0.3	2.99±0.01	2.50±0.99	7.5±3.0
CTPR3ΔS	3.17±0.01	2.65±0.08	8.4±0.3	3.17±0.01	2.71±0.13	8.6±0.4
CTPR3	3.36±0.01	2.89±0.13	9.7±0.4	3.29±0.01	3.61±0.13	11.9±0.4
CTPRa2ΔA	1.01±0.04	1.74±0.05	1.8±0.1			
CTPRa2ΔS	1.63±0.02	1.87±0.07	3.0±0.1			
CTPRa2	1.71±0.04	2.07±0.19	3.5±0.3			
CTPRa3ΔA	2.18±0.01	2.34±0.08	5.1±0.2			
CTPRa3ΔS	2.31±0.01	2.30±0.09	5.3±0.2			
CTPRa3	2.39±0.02	2.42±0.12	5.8±0.3			

Errors are calculated from the fit of the graph. [a] Calculated from Equation 4.2.
[b] Calculated from Equation 4.3

% more solvent exposed than the next most solvent exposed helix. Interestingly, the differing stabilities of the N-terminal and C-terminal helices explains why crystal structures of certain CTPRs show the S helix being preferentially unfolded (missing density) and its position occupied by an N-terminal A helix from another CTPR protein [72] .

4.3.5.1.4 Chemical Denaturation in MOPS buffer

Recently, it was shown that MOPS is more tolerant to changes in salt concentration than phosphate buffers at pH 7 [113]. The consequence of this is that as the GuHCl concentration increases, the pH of the phosphate buffered samples decreases. This lowers the pH of the sample closer to the pI of the protein (approximately 4 for CTPR proteins). Denaturations in this thesis were performed in phosphate to allow comparison with published data. Also phosphate is known to stabilise proteins, therefore we wanted to confirm that the changes seen the data was not a consequence of this effect. Therefore, the denaturations were repeated in MOPS buffer for the CTPRn series. Figure 4.7 shows that the proteins display the same trends as that

seen for the phosphate buffered samples, where the mid-point increases with repeat number and that the ΔA mutants are less stable than the ΔS mutants. The samples were prepared in the same manner as for the phosphate buffered samples except 50 mM MOPS at pH 7 was used as the buffer. Measurements were recorded with the samples in a 5 mm pathlength cuvette held at 10 °C in a thermostated cuvette holder.

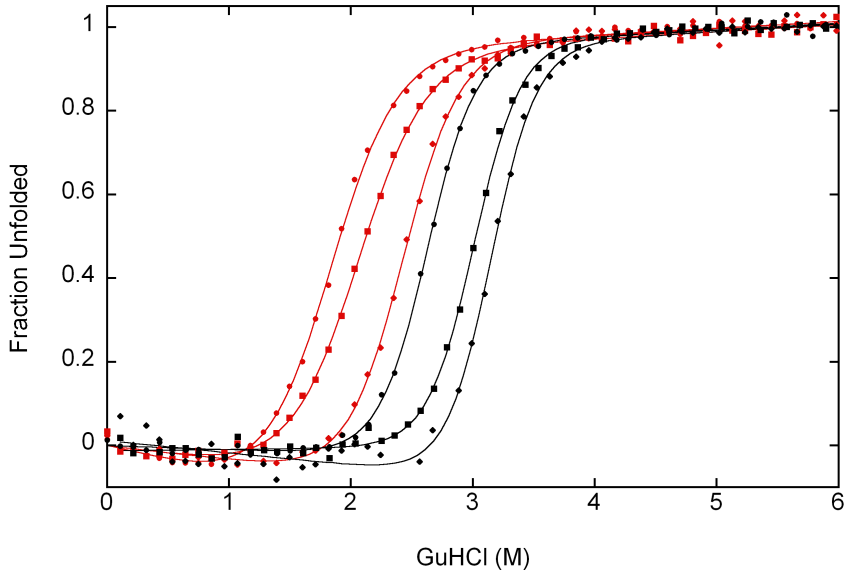

Figure 4.7: Normalised chemical denaturation curves for the CTPRn series CTPR2 (red diamonds), CTPR2ΔS (red squares), CTPR2ΔA (red circles), CTPR3 (black diamonds), CTPR3ΔS (black squares) and CTPR3ΔA (black circles). All proteins undergo a single reversible transition from the folded to unfolded state. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM MOPS pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

When compared to the phosphate data for the same proteins we can see a decrease in the mid-points of denaturation for all the proteins studied. The decrease is consistently around 0.2 M. This shows that there is a stabilising effect from the change in pH that would be observed when the denaturation is conducted in MOPS instead of phosphate. However the same trend in stabilities was observed, the ΔA mutation was more destabilising than the ΔS mutation.

Table 4.2: Thermodynamic parameters for the GuHCl induced unfolding of CTPR proteins in MOPS buffer

Construct	**MOPS Data Set**			**Phosphate Data Set 1**		
	[a] $[D]_{50\%}$ (M)	[a] m-value kcalmol^{-1}M^{-1}	[b] $\Delta G^{H_2O}_{D-N}$ kcalmol^{-1}	[a] $[D]_{50\%}$ (M)	[a] m-value kcalmol^{-1}M^{-1}	[b] $\Delta G^{H_2O}_{D-N}$ kcalmol^{-1}
CTPR2ΔA	1.81±0.01	2.09±0.05	3.8±0.1	1.94±0.01	2.08±0.05	4.0±0.1
CTPR2ΔS	2.04±0.01	2.01±0.05	4.1±0.1	2.21±0.01	2.10±0.03	4.6±0.1
CTPR2	2.42±0.01	2.44±0.07	5.9±0.2	2.57±0.01	2.43±0.06	6.2±0.2
CTPR3ΔA	2.67±0.01	2.78±0.07	7.4±0.2	2.97±0.01	2.82±0.10	8.4±0.3
CTPR3ΔS	3.07±0.01	2.96±0.11	9.1±0.3	3.17±0.01	2.65±0.08	8.4±0.3
CTPR3	3.14±0.01	3.00±0.17	9.4±0.5	3.36±0.01	2.89±0.13	9.7±0.4

Errors are calculated from the fit of the graph. [a] Calculated from Equation 4.2.
[b] Calculated from Equation 4.3

4.3.6 The Ising Model

The Ising model was developed by Ernst Ising to describe the interactions between magnetic dipoles [114]. Since then the model has been adapted and used to describe a variety of systems that involve a linear array of interacting subunits, such as phase transition within polymers [115].

4.3.6.1 Homozipper Ising Model

The modularity of repeat proteins means that their equilibrium unfolding has been analysed using a 1D homozipper Ising model [9, 73, 105, 116–118]. This models each folding unit (helix or repeat depending on the complexity of the model) as an independent folding motif that can be either folded or unfolded. Once the unit is folded its nearest unfolded neighbour can gain favourable interactions by folding.

Homozipper Ising models treat each element of a repeat protein as an identical individual folding unit, with the same intrinsic and interfacial energy. This captures many aspects of the equilibrium unfolding and can successfully predict the denaturation curves of other ensembles of the folding units. However, it cannot adequately describe the unequal distribution of stabilities, that results from the different types of helices that form a CTPR.

4.3.6.2 Heteropolymer Ising Model

By creating a set of deletion mutants lacking either the N-terminal A helix or C-terminal S helix, it was possible to construct a more complex Heteropolymer Ising model that differentiates between the contributions to stability by the N-terminal A helix, C-terminal S helix and the helices that make up the rest of the protein [117]. This allows us to calculate/simulate any 'switch-like functionality' (S-switch) e.g.

the population of a denatured C-terminal S-helix with the rest of the protein folded relative to any other unfolding (within each protein) throughout the chemical unfolding transition.

The Heteropolymer Ising was constructed of a linear algebraic series of equilibrium constants that represent the intrinsic folding stability and the interfacial energy terms of each helix in a nearest-neighbour array (ΔG_i and $\Delta G_{i-1,i}$ respectively). In our model, 9 variables delineate the contributions made by the N-terminal capping A-helix (ΔG_i^A, $\Delta G_{i-1,i}^A$ and m^A), internal I-helices (ΔG_i^I, $\Delta G_{i-1,i}^I$ and m^I) and the C-terminal capping S-helix (ΔG_i^S, $\Delta G_{i-1,i}^S$ and m^S). The m parameters gave a denaturant dependence to the intrinsic stabilities (m^A, m^I and m^S).

The expressions defining the equilibrium constants (Equations 4.4 and 4.5) are given below:

$$\kappa_i = e^{-(G_i+(mx))/RT} \tag{4.4}$$

Where G_i is the free energy of folding for the helix at position i, with denaturant sensitivity m and at denaturant concentration x. R is the gas constant and T is experimental temperature.

$$\tau_{i-1,i} = e^{[-G_{i-1,i}/RT]} \tag{4.5}$$

Where $G_{i-1,i}$ is the free energy for the interface between helices at positions $i-1$ and i. R is the gas constant and T is experimental temperature. The partition function, $q(n)$, is given by Equation 4.6.

$$\mathrm{q}(n) = [1\ 0] \begin{bmatrix} \kappa_1^A & \upsilon_1 \\ \kappa_1^A & \upsilon_1 \end{bmatrix} \begin{bmatrix} \kappa_2^B \cdot \tau_{1,2}^B & \upsilon_2 \\ \kappa_2^B & \upsilon_2 \end{bmatrix} \left\{ \begin{bmatrix} \kappa_i^A \cdot \tau_{i-1,i}^A & \upsilon_i \\ \kappa_i^A & \upsilon_i \end{bmatrix} \begin{bmatrix} \kappa_j^B \cdot \tau_{i,j}^B & \upsilon_j \\ \kappa_j^B & \upsilon_j \end{bmatrix} \right\}^{\left(\frac{n-3}{2}\right)} \begin{bmatrix} \kappa_n^S \cdot \tau_{n-1,n}^S & \upsilon_n \\ \kappa_n^S & \upsilon_n \end{bmatrix} [1\ 1] \tag{4.6}$$

The full partition function of the protein with n helices is given by q(n), where all $\upsilon = 1$. This defines the fully-folded state. The model allows for fitting of separate

parameters (κ and τ, thus G_i, $G_{i-1,i}$and m) to describe behaviour of the A, B and S-helices by globally fitting to data for degenerate CTPR protein compositions. A subpartition function, q(i), must be calculated by considering only the folded state of each helix in turn, given by iteratively parsing the value $\upsilon = 0$ to the term for each helix. The fraction folded, θ_F is then simply defined as the sum of the subpartition functions divided by the number of terms (helices) multiplied by the full partition function, Equation 4.7.

$$\theta_F = \sum_{n}^{i=1} \frac{q\,(i)}{n.q\,(n)} \tag{4.7}$$

A numerical solution of the Ising model was determined by globally fitting the chemical denaturation of each CTPR series. This consisted of fitting 12 denaturation curves of the CTPRn series (6 proteins with duplicated curves) and 10 denaturation curves of the CTPRan series, Figure 4.8 (CTPRa2ΔA removed due to lack of native baseline and denaturations of CTPRa2, CTPRa3, CTPRa4, CTPRa5, CTPRa6, CTPRa8 and CTPRa10 were completed by Dr Javadi).

A single global minimum was ensured by seeding 1000 random searches, each with 1000 trajectories, using the Mathoptimizer module of Mathematica (Wolfram). Our model fit well to the experimental data with a root mean square of residuals from the fit to CD data for the series of CTPRan was 0.44 and for the series of CTPRn was 0.38, equivalent to <2 % of the data amplitude.

Table 4.3 lists the values obtained for the fitted parameters. These show that the interfacial energies were all stabilising ($\Delta G_{i-1,i}^{H_2<0}$) and the intrinsic stabilities were not ($\Delta G_i^{H_2O<0}$). From the fitted variables the stability of any TPR ensemble ($\Delta G_{0\rightarrow j}^{H_2O<0}$) can be calculated by simply adding energy terms.

Thus, to obtain a folded CTPR protein the additive effects of the favourable stabilising interfaces only outweighed the energetically unfavourable intrinsic helix sta-

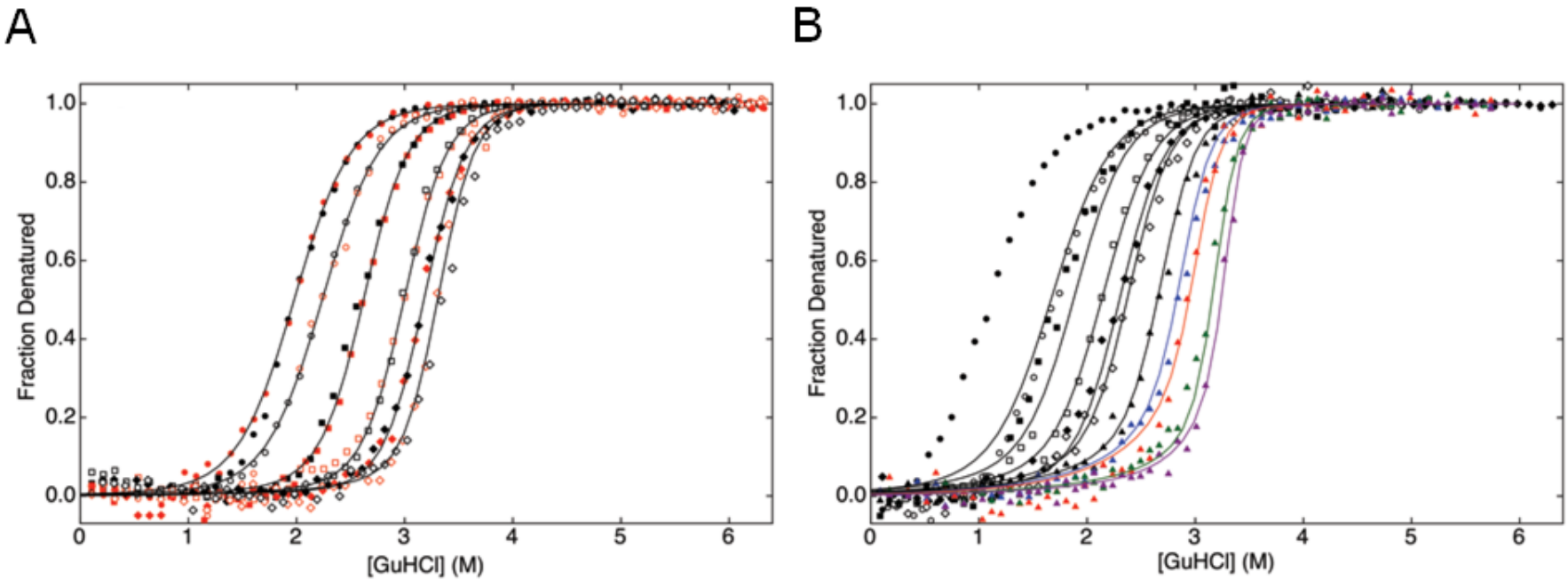

Figure 4.8: Chemical Denaturations for the CTPRn (A) and the CTPRan (B) series globally fit to the Heteropolymer Ising model

bilities when ensembles of more than two helices are combined. This is in agreement with previous experimental data and data analysed with the homozipper Ising model [9,105]. Our new data and analysis can also discriminate between the differing energetic contributions of N-terminal A-helix, C-terminal S-helix and the internal repeat helices.

In agreement with the previous sections qualitative analysis, these show that the removal of the N-terminal A-helix is more destabilising than removing an internal helix or C-terminal S-helix. Interestingly, although the differences between helices are not large in magnitude, they become more pronounced with addition of GuHCl due to their differing m-values.

Table 4.3: Values obtained from the fitting of the equilibrium unfolding data to the Heteropolymer Ising model

CTPR Series	**N-terminal Capping α-helix (A)**				**Internal α-helices (I)**				**C-terminal Capping α-helix (S)**			
	ΔG_i^A kcalmol^{-1}	m_A kcalmol^{-1}M^{-1}	$\Delta G_{i-1,i}^{H_2O}$ kcalmol^{-1}	$\Delta G_{0\rightarrow1}^{H_2O}$ kcalmol^{-1}	ΔG_i^I kcalmol^{-1}	m_I kcalmol^{-1}M^{-1}	$\Delta G_{i-1,i}^I$ kcalmol^{-1}	$\Delta G_{0\rightarrow1}^{H_2O}$ kcalmol^{-1}	ΔG_i^S kcalmol^{-1}	m_S kcalmol^{-1}M^{-1}	$\Delta G_{i-1,i}^S$ kcalmol^{-1}	$\Delta G_{0\rightarrow1}^{H_2O}$ kcalmol^{-1}
CTPRn	5.7	0.6	-8.8	-3.1	3.5	0.5	-6.1	-2.6	0	0.8	-2.9	-2.9
CTPRan	5.6	0.85	-8.2	-2.6	2.3	0.5	-4	-1.7	0.4	0.7	-2	-1.6

$\Delta G_{0\rightarrow1}^{H_2O} = \Delta G_i^{helix in H_2O} + \Delta G_{i-1,i}^{helix in H_2O}$ The stability gained when a single helix is added to a folded TPR ensemble. The root mean square of residuals from the fit to CD data for the series of CTPRan was 0.44 and for the series of CTPRn was 0.38, equivalent to <2 % of the data amplitude. Random seeding with iterative convergence by steepest descents and simulated annealing gave the same values within one same values within one decimal place, indicating that global minima were reliably obtained.

4.3.6.3 Modelling of a Thermodynamic Switch in CTPR3

A 'good/efficient' switch is a product of a protein with a high overall stability and with a suitable differential in stability between capping and internal helices. To test this hypothesis and the Ising model's predictive qualities it was logical to redesign and model a chimera that combined the more stable core of the smaller CTPR3 from CTPRn series with the less stable final S-helix of the CTPRan series (i.e. inserting the `PNN` to `PRS` mutation). Figure 4.9A shows a schematic of CTPRy3 with its C-terminal helix unfolded.

The new construct was termed CTPRy3 and its chemical denaturation and S-switch function was simulated using the Ising model with the experimentally fitted values for the A and I helices of CTPRn series and S-helix from the CTPRan series. The model predicted that CTPRy3 would have a maximum S-switch of $\approx .\ 50$ % and further helical unfolding of only $\approx .\ 15$ % at 2.8 M GuHCl, Figure 4.9.

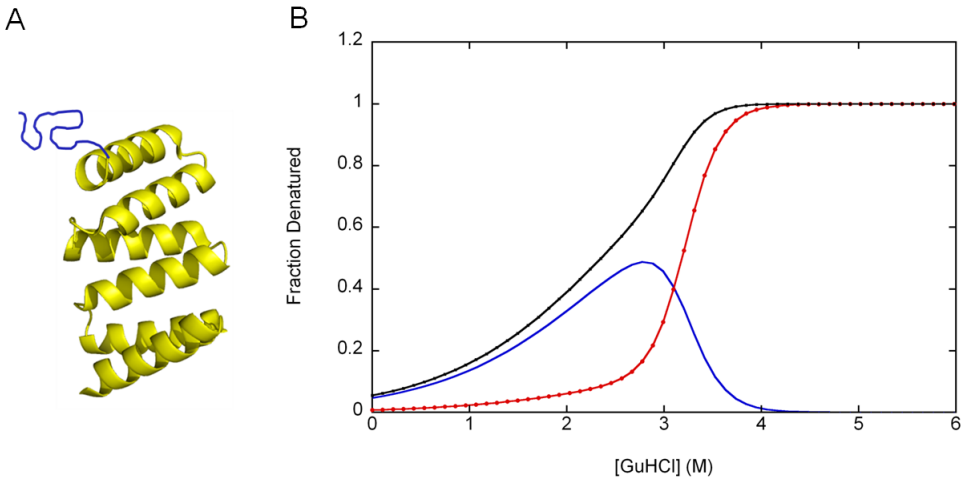

Figure 4.9: (A) Proposed intermediate of CTPRy3, where the C-terminal helix is unfolded (shown as a random coil in blue) leaving the rest of the protein folded (constructed from 1NA0). (B) Simulated data from the Heteropolymer Ising model. The population of CTPRy3 with the C-terminal helix unfolded as a function of GuHCl concentration (blue line). Simulation of the chemical denaturation curve of CTPRy3 (red line) and the fractional population of CTPRy3 that have the C-terminal helix unfolded (black line). Interestingly, you can see that 90 % of the population of CTPRy3 has the C-terminal helix unfolded by the mid-point of the denaturation curve.

4.3.7 Summary of Equilibrium unfolding characteristics of the CTPRs

By studying the stability of a range of deletion mutants, we were able to determine that removing the N-terminal helix caused a greater reduction in stability than removing the C-terminal helix. Both series of CTPR displayed this trend. Due to the smaller contribution to the stability of the protein, the C-terminal helix is the more suitable target for redesign as a switch. We know that the `PNN` to `PRS` mutation results in a decrease in the stability of the protein [9]. Conveniently, this is situated in the loop region before the C-terminal helix. Thus constructing the chimera of the two CTPR series (CTPRn core and CTPRan C-terminal sequence) we can create a large enough difference in stabilities between the core of the protein and the C-terminal helix to allow access to a partially populated state. A Heteropolymer Ising model was used to simulate the fractional population with the C-terminal helix unfolded at a given GuHCl concentration would be. According to the results of the simulation this switching protein, CTPRy3, should give us access to a C-terminal helix unfolded intermediate, at a GuHCl concentration of 2.8 M.

4.3.8 CTPRy3

Following the development of the Heteropolymer Ising model, the proposed 'S-switch' protein, CTPRy3 was produced and it's stability and ability to 'switch' was assessed, using chemical and thermal denaturation and 2D NMR. Figure 4.10 shows the site of the NN→RS mutation in the crystal structure of CTPR3.

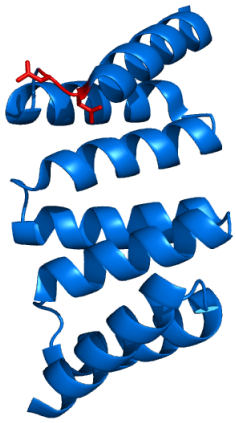

Figure 4.10: Crystal structure of CTPR3 (1NA0). Highlighted in green are the two asparagine residues that are to be mutated to arginine and serine to create the mutant CTPRy3.

4.3.8.1 Cloning, Expression and Purification of CTPRy3

CTPRy3 was cloned as described in Section 2.1 and Table 2.1 and the mutations were confirmed via DNA sequencing. The protein was expressed and purified under native conditions and purified as described in Chapter 2.3.2 and gave high yields of approximately 10 mg/L of culture.

4.3.8.2 Characterisation of CTPRy3

CTPRy3 was analysed using size exclusion chromatography to confirm that it is monomeric. A 100 μl sample at a concentration of 50 μM was loaded onto a GE Healthcare Superdex 75pg 300/10 column (equilibrated with 50 mM phosphate pH

7, 150 mM NaCl). The flow rate was set at 0.5 ml/minute. The elution of the protein was monitored using UV absorbance at 280 nm. The chromatogram for CTPRy3 was compared with that of CTPR3 (Figure 4.11A). The mutant was confirmed as monomeric.

Far-UV CD wavelength scans of CTPRy3 were performed to assess the secondary structure, Figure 4.11 B. The wavelength scan was performed on a 5 μM sample, dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 ^{o}C. The scans gave a single negative peak at 222 nm showing that CTPRy3 is α-helical. The values for the molar ellipticity (expressed as mdeg $cm^{-2}dmol^{-1}$) were CTPR3 -2.44x10^5 and CTPRy3 -2.35x10^5. As the values are so close the CTPRy3 construct has the same helical content as CTPR3 and thus is correctly folded.

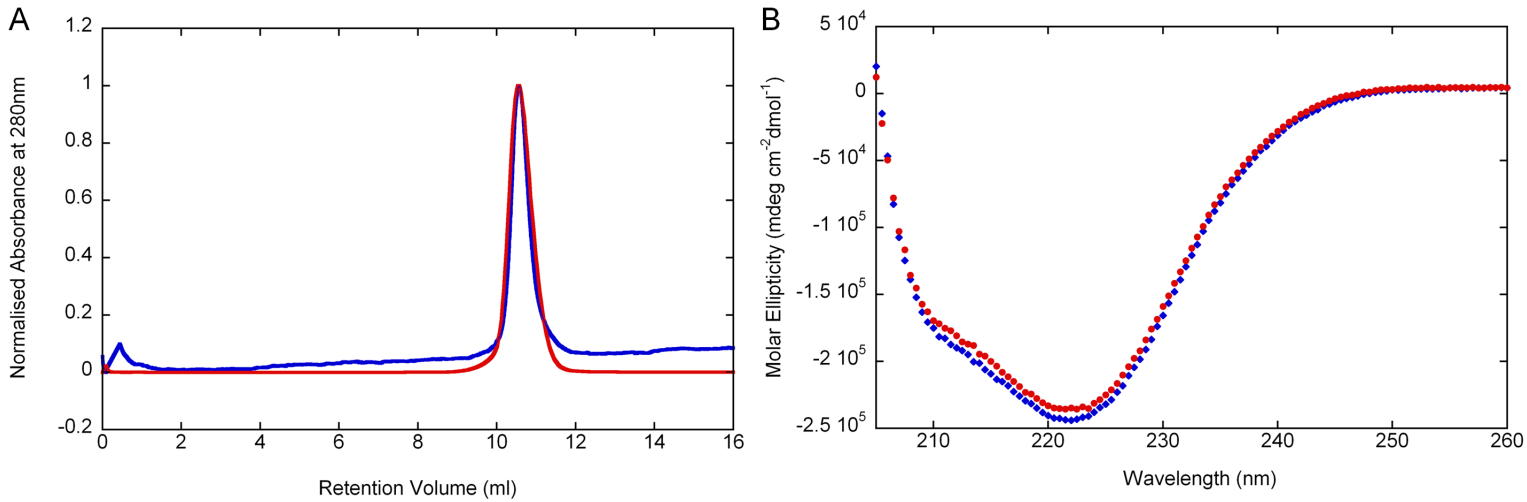

Figure 4.11: (A) Analytical size exclusion chromatograms of CTPRy3 (blue) and CTPR3 (red). Conditions were: 50 mM phosphate pH 7, 150 mM NaCl, buffer flow rate was set at 0.5 ml/minute. The column used was a GE Healthcare Superdex 75pg 300/10. The elution of the protein was monitored using the absorbance at 280 nm and the retention on the column measured in ml. The retention time is the same for both proteins, this shows that CTPRy3 is monomeric. (B) Far UV wavelength scan of CTPR3 (blue) and CTPRy3 (red) at 5 μM dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. The scan of CTPRy3 showed a single negative peak at 222 nm, indicating that the protein is highly α-helical.

4.3.8.3 Chemical Denaturation of CTPRy3

The stability of CTPRy3 was assessed via GuHCl denaturations followed using Circular Dichroism at 222 nm (Figure 4.12). The GuHCl denaturations were performed as described in Section 2.5.2.2. The denaturations were performed in 50 mM phosphate pH 7. The samples were held in a 5 mm pathlength cuvette held at 10 °C in a thermostated cuvette holder. CTPRy3 underwent a single transition from the folded to the unfolded state. The mid-point of denaturation was 3.2 M ± 0.01 M, for all 3 experiments.

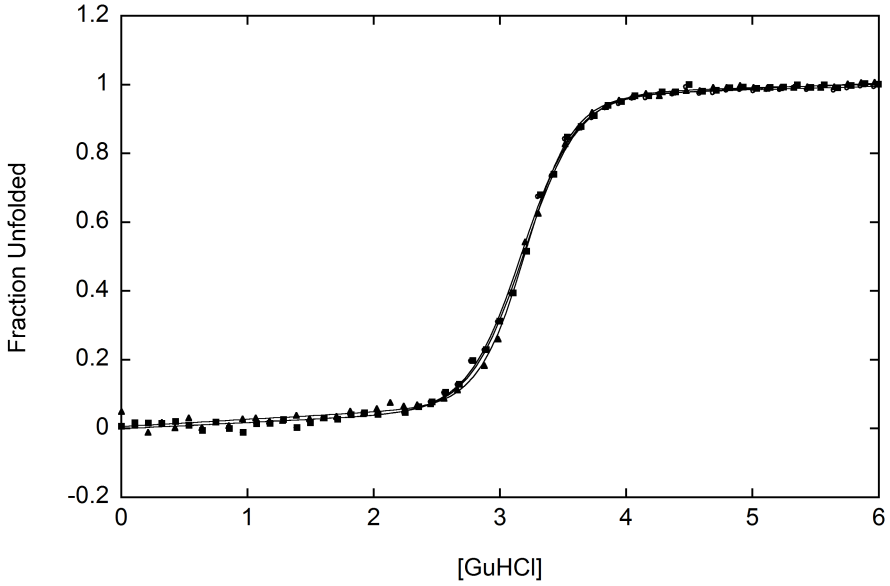

Figure 4.12: Normalised chemical denaturation curves for CTPRy3, there are 3 data sets. All data sets show that the protein undergoes a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baselines equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation that arises in concentration between data sets. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. GuHCl was used to denature the protein and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

4.3.8.3.1 Comparison of the Chemical Denaturations of CTPRy3 with the Simulated Data from the Ising Model

When the data from the chemical denaturations of CTPRy3 was compared with the simulated data from the Heteropolymer Ising model, the simulation is almost indistinguishable from the experimental data, Figure 4.13. This shows that the Heteropolymer Ising model can accurately predict the denaturation curves of CTPR proteins.

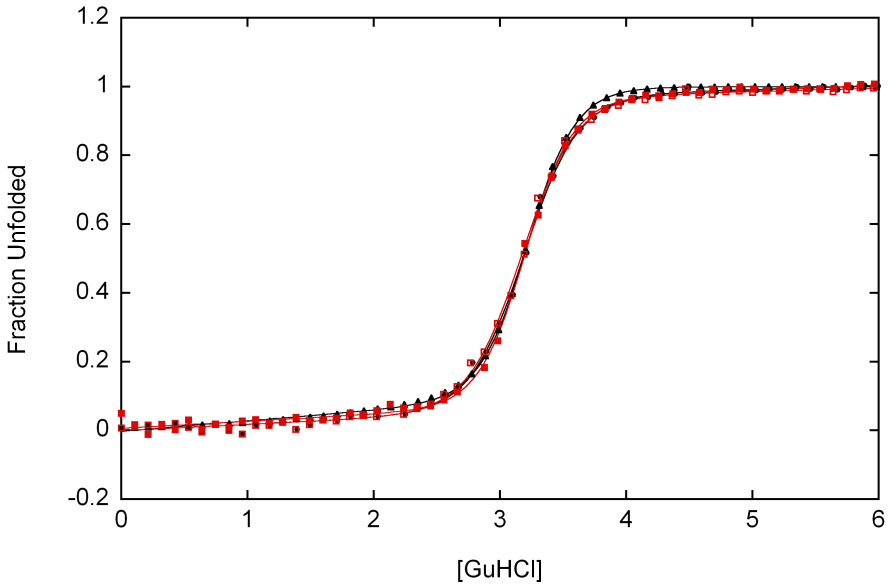

Figure 4.13: Normalised chemical denaturation curves for CTPRy3 (from Figure 4.12) in red overlaid with the Ising model simulated data for the same construct in black. The simulated data is almost indistinguishable from the experimental data.

4.3.8.3.2 Comparison of Chemical Denaturations of CTPRy3 with CTPR3 and CTPR3ΔS

The denaturation of CTPRy3 showed a single transition from the folded to the unfolded state, Figure 4.14 A. As predicted, this transition has a lower midpoint than CTPR3, showing it is less stable. The data for CTPRy3 can be qualitatively compared to data for the other CTPR3 mutants by fitting to a two-state two-sloping baselines equation 4.2. The mid-points calculated from the fit of the data can be

used to compare the stability of the different proteins.

Interestingly, the mid-point of CTPRy3 is the same as that seen for CTPR3ΔS, indicating that, at this concentration of GuHCl, the CTPRy3 has the same helical content as CTPR3ΔS. Moreover, there is an observable change in the slope of CTPRy3's folded baseline in comparison to CTPR3, Figure 4.14 B. The baselines for CTPR3 and CTPR3ΔS are flat, with almost a negative slope. This is in line with what is observed for CTPR proteins. However, the baseline of CTPRy3 has a noticeable slope before the first transition. This deviance from the trend that is characteristic for the CTPR proteins is very significant and indicates that there is a variation the structural changes that the CTPRy3 undergoes when compared with the similar CTPR3 proteins. When the ellipticity was corrected for concentration, the maximum difference in baselines between the two proteins occurs at approximately 2 M GuHCl and equates to 7 % of the signal at 222 nm. The complete loss of the C-terminal helix would equate to a 14 % reduction in signal.

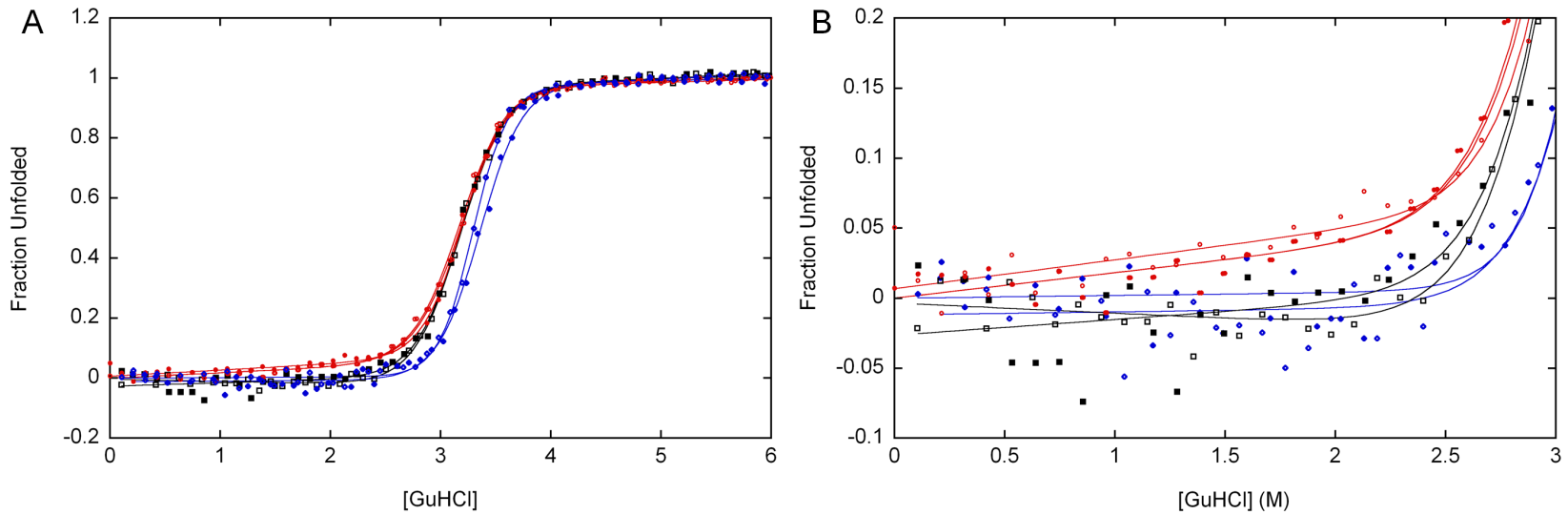

Figure 4.14: (A) Normalised chemical denaturation curves of: CTPR3 (blue), CTPR3ΔS (black) and CTPRy3 (red). (B) Shows an enlargement of the folded baseline, showing that the baseline of CTPRy3 has a much steeper gradient than either CTPR3 or CTPR3ΔS. All the proteins undergo a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baseline equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation in ellipticity that arises from the different number of helices for each construct. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

4.3.8.3.3 Chemical Denaturation of CTPRy3 in MOPS

Previously we saw that changing the buffer system from phosphate to MOPS caused an decrease in the stability of the proteins studied. To confirm that this same trend was seen with CTPRy3, the denaturation was repeated in 50 mM MOPS pH 7, Figure 4.15.

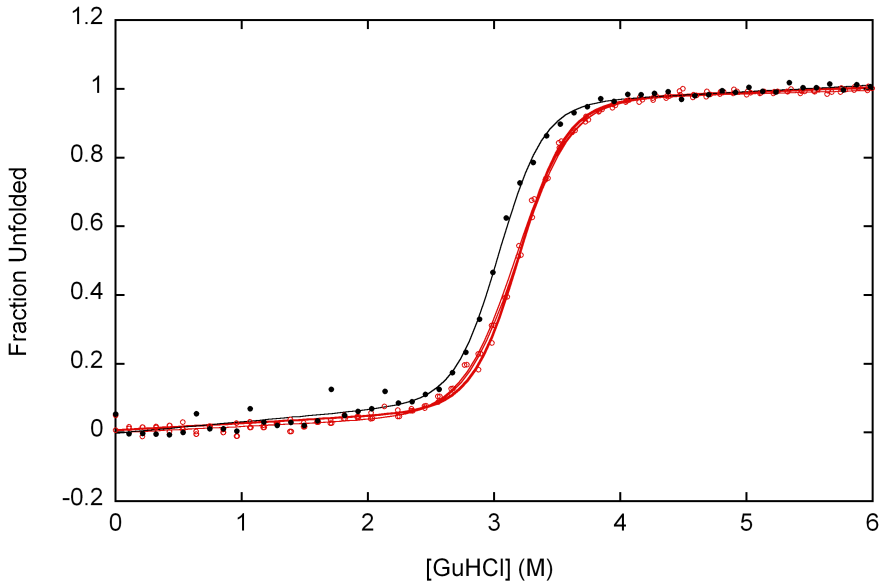

Figure 4.15: Normalised chemical denaturation curves for CTPRy3, in 50 mM MOPS pH 7 (black) and in 50 mM phosphate pH 7 (red). All data sets show that the protein undergoes a single reversible transition from the folded to unfolded state. The MOPS denaturation shows the same transition, but has a lower mid-point. This shift of approximately 0.15 M GuHCl is in agreement with the changes observed with the other CTPR proteins. Initially, the data was fit using a two-state two-sloping baseline equation (using Equation 4.1). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation that arises in concentration between data sets. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in either 50 mM MOPS pH 7 or 50 mM phosphate pH 7. GuHCl was used to denature the protein and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 ^{o}C.

4.3.8.3.4 Comparison of CD Wavelength Scans of CTPRy3, CTPR3 and CTPR3ΔS

It was evident from the CD denaturation unfolding curves that there the CTPRy3 (in comparison to CTPR3 and CTPR3ΔS) undergoes a different structural change in the region before transition. To further investigate this, CD wavelength scans were recorded for CTPRy3, CTPR3 and CTPR3ΔS, with and without 2 M GuHCl, Figure 4.16. These show that in 0 M GuHCl, the ellipticity of CTPRy3 is the same as CTPR3, indicating that they have the same helical content. The molar ellipticity (expressed as mdeg $cm^{-2}dmol^{-1}$) at 222 nm in 0 M GuHCl was: CTPR3 $-2.4x10^5$; CTPRy3 $-2.3x10^5$ and CTPR3ΔS $-2.1x10^5$ The molar ellipticity (expressed as mdeg $cm^{-2}dmol^{-1}$) at 222 nm in 2 M GuHCl was: CTPR3 $-2.5x10^5$; CTPRy3 $-2.2x10^5$ and CTPR3ΔS $-2.2x10^5$. However at 2 M GuHCl, the ellipticity of CTPRy3 has reduced to the same as that seen for CTPR3ΔS. At 2 M GuHCl the helical content of CTPRy3 is equal to that of CTPR3ΔS, indicating that, the C-terminal helix of CTPRy3 is unfolding.

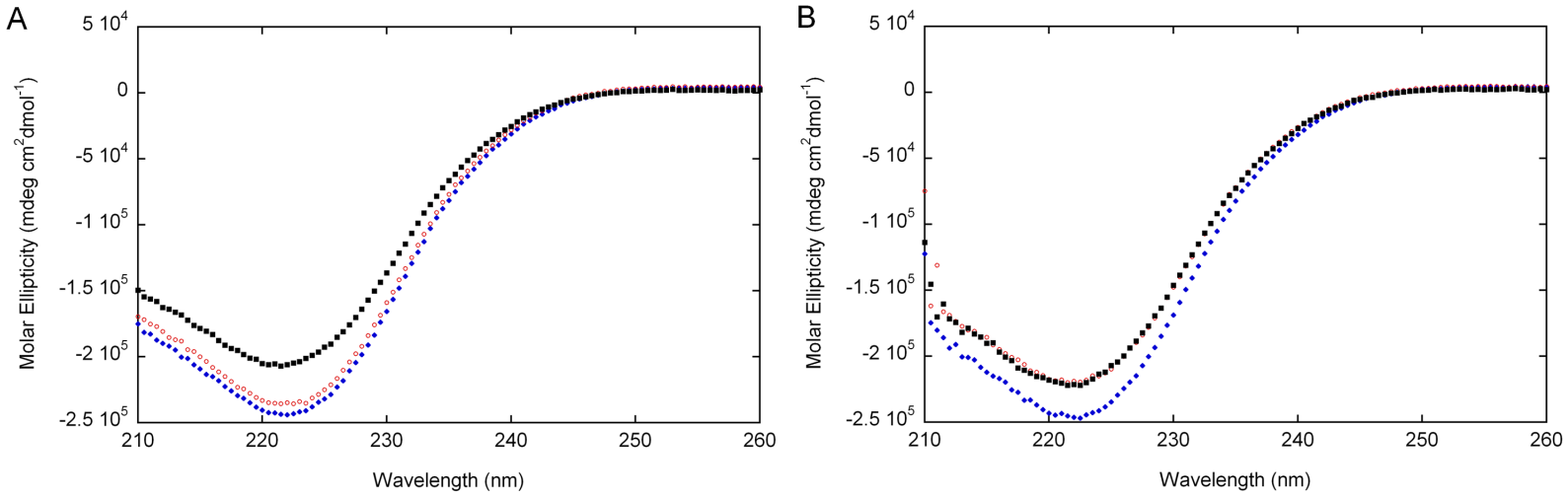

Figure 4.16: Far UV CD wavelength scans of CTPR3 (blue), CTPR3ΔS (black) and CTPRy3 (red) at 0 M GuHCl (A) and 2 M GuHCl (B). The protein concentration is 5 μM and the buffer was 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. At 0 M GuHCl the molar ellipticity of CTPRy3 is equal to that of CTPR3 However, at 2 M the molar ellipticity of CTPRy3 is the same as CTPR3ΔS.

4.3.8.4 2D NMR HSQC spectra of CTPRy3 and CTPR3

To determine whether the change in α-helical signal observed in CD spectra at low GuHCl concentrations was due to the unfolding of CTPRy3's C-terminal α helix, 2D NMR HSQC experiments were performed for both N^{15} CTPRy3 and N^{15} CTPR3.

N^{15} labelled CTPRy3 and CTPR3 were expressed and purified as described in Section 2.3.7.2. Figure 4.17 shows an overlay of the spectra for CTPR3 and CTPRy3, where there appears to be very little change in the number and position of the cross-peaks. Samples were prepared for HSQC experiments, by dialysing the protein into 50 mM phosphate pH 6.8, 150 mM NaCl. The samples were concentrated to 400-800 μM. 10% (v/v) of D_2O was added. These conditions were used to facilitate comparison with published data, selecting a different buffer system could have significantly changed the chemical shifts. Data was collected at 25 °C using a Bruker AV600, 600MHz spectrophotometer.

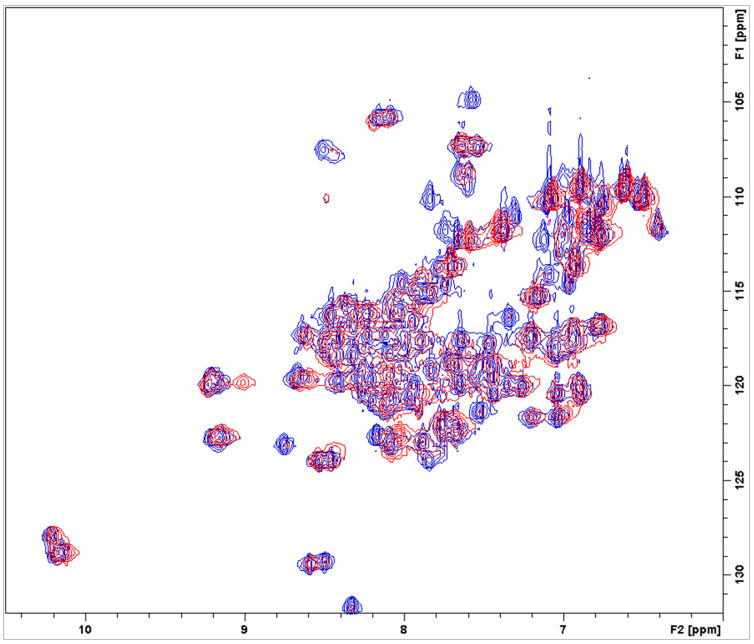

Figure 4.17: 2D HSQC NMR spectra of CTPR3 (blue) and CTPRy3 (red). Buffer conditions were 50 mM phosphate pH 6.8, 150 mM NaCl, with 10% (v/v) of D_2O.

4.3.8.4.1 Assignment of CTPRy3 and CTPR3 spectra

Spectra were processed using TopSpin 2.1 patch level 6. Cross-peaks for CTPR3 were assigned by comparing spectra collected for CTPR3 with published assigned data. Within CTPR3 there are 103 residues with assigned peaks out of a possible 117. There are 120 residues in total in CTPR3 but 3 of these are proline residues that don't have an amide proton and are therefore not visible to 2D NMR.

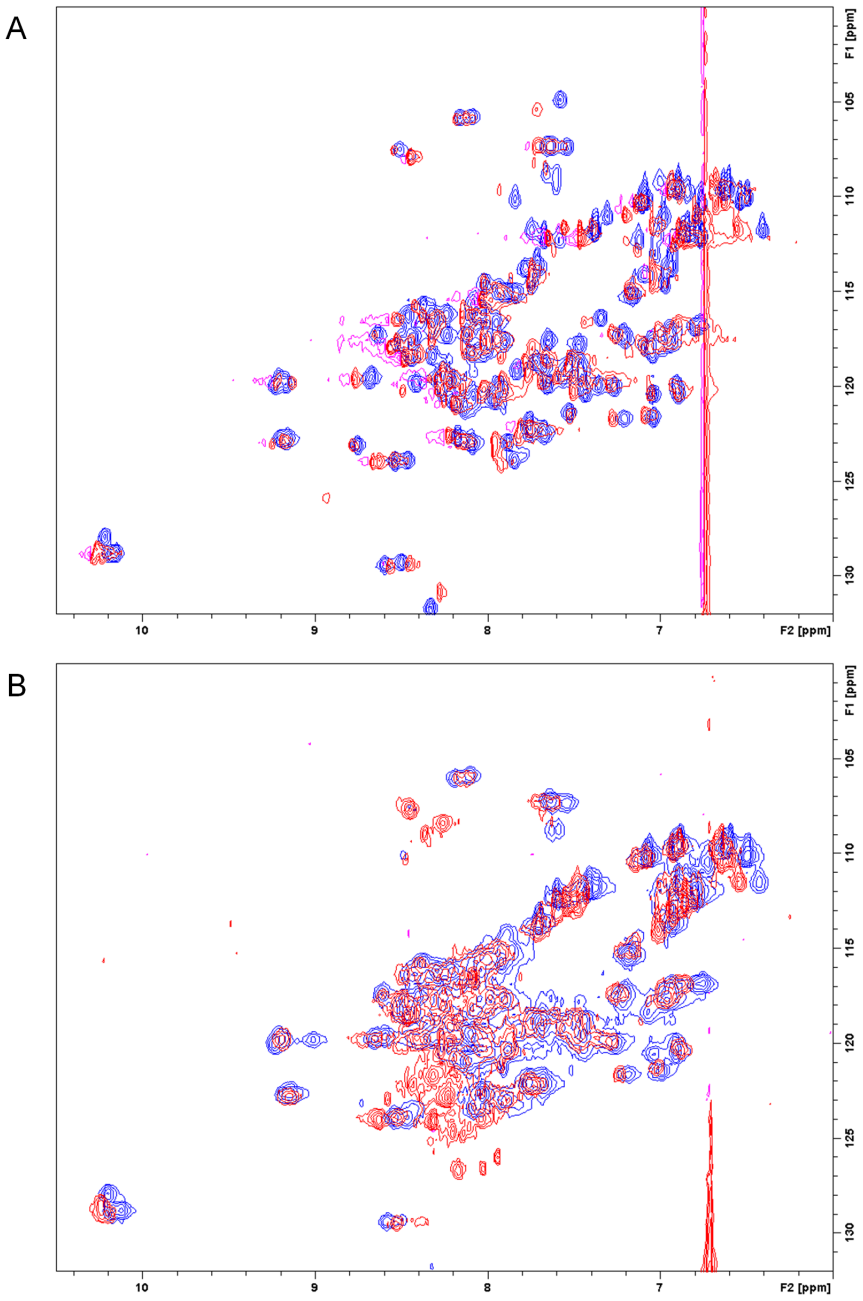

Figure 4.18: (A) 2D NMR HSQC data for CTPR3 in 0 M GuHCl (blue) and 1.6 M GuHCl (red). (B) 2D NMR HSQC data for CTPRy3 in 0 M GuHCl (blue) and 1.6 M GuHCl (red). Buffer conditions were the same for both samples: 50 mM phosphate pH 6.8, 150 mM NaCl. Data was recorded at room temperature.

CTPRy3 was assigned by comparing CTPRy3 with the CTPR3 data. A total of 91 peaks were able to be assigned in CTPRy3. Figure 4.17 shows an overlay of the two spectra for CTPR3 and CTPRy3 at 0 M GuHCl. There is very little difference between the two spectra confirming that CTPRy3 adopts the desired

CTPR fold. The proteins were partially denatured with GuHCl and the 2D NMR HSQC recorded again. The assignment of the partially denatured data (red) was accomplished through comparison with the 0 M sample (blue). When 1.6 M GuHCl was added to CTPR3, the spectra showed a loss of 18 peaks, but overall the protein remained folded. Figure 4.18A, shows the 0 M and 1.6 M GuHCl spectra. The data for CTPRy3 at 1.6 M shows that there were 27 less peaks than in the 0 M spectra, Figure 4.18B.

4.3.8.4.2 Comparison of CTPR3 and CTPRy3 2D NMR HSQC spectra

In 0 M GuHCl, each protein gave well dispersed spectra, with little perturbation in chemical shifts between the amide protons of CTPR3 and CTPRy3. Unique probes were therefore available throughout CTPRy3 structure, specifically 12 residues out of 15 in the C-terminal helix, to report on the dynamics and stability on per a residue basis.

Initially both CTPR3 and CTPRy3 proteins were incubated with 2 M GuHCl and the HSQC experiments repeated. These GuHCl concentrations were chosen as the Heteropolymer Ising model predicted a 50% population of S-helix unfolded intermediate at 2 M GuHCl (higher GuHCl concentrations were not used as the model predicted that further helices of CTPRy3 would unfold). As predicted CTPR3 undergoes little change between 0 and 2 M GuHCl; however this concentration resulted in too great a change in the CTPRy3 to allow full assignment of the spectra. Consequently, the experiments were repeated with a lower concentration of GuHCl (1.6 M). Here we observed a significant and specific change in the spectra of CTPRy3.

Upon addition of 1.6 M GuHCl 17 % of cross-peaks were lost overall in the CTPR3 (18 of 103 peaks) spectrum and 30 % were lost from the CTPRy3 (27 out of 91

peaks). When we look specifically at the cross-peaks relating to the C-terminal helix: we see that the CTPR3 lost 36 % of cross-peaks (5 out of 14 peaks), whereas the CTPRy3 lost 83 % (10 out of 12). The number of cross-peaks that are lost from the main body (i.e. a CTPR3ΔS unit) from CTPR3 was 12 out of 89 and the number lost from the CTPRy3 spectra was 17 out of 79. When expressed as a percentage there is little change for either protein, CTPR3 loses 13 % and CTPRy3 loses 22 %, indicating that the main body remains folded.

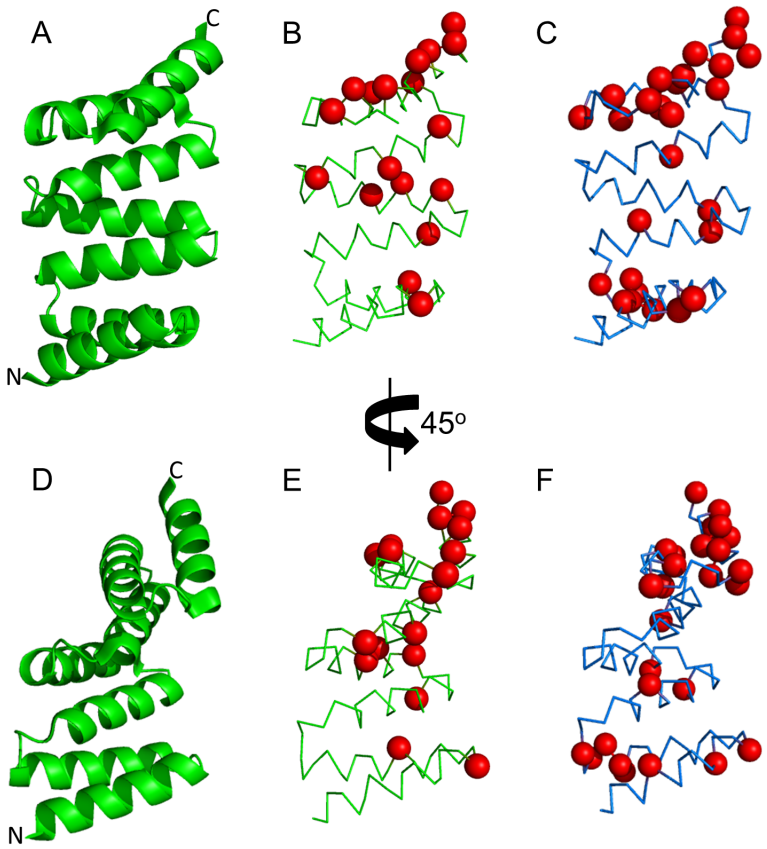

Figure 4.19: Changes in the HSQC spectra (between 0 M and 1.6 M GuHCl) mapped onto the structure of CTPR3 (1NA0) (A and D). Cross-peaks that are absent from the 1.6 M spectra are marked using a red sphere. CTPR3 (B and E) undergoes less structural changes as shown by the smaller loss of peaks. CTPRy3 (C and F) has more significant changes in the structure and this is very strongly seen in the loss of 10 of the 12 peaks from the C-terminus.

Chemical shift perturbations (CSP) values were used to quantify the changes observed for individual cross-peaks between the two spectra. For each cross-peak, the change in chemical shift for both the N^{15} and the H^{1} spectra is calculated. These

values are used in Equation 4.8 [119] to calculate the CSP values.

$$CSP = \sqrt{\left(\frac{\delta \mathrm{N}}{5}\right)^2 + \left(\frac{\delta \mathrm{H}}{2}\right)^2} \tag{4.8}$$

δN is the difference in the chemical shift for an assigned cross-peak between the two N^{15} spectra. δH is the difference in the chemical shift for an assigned cross-peak between the two H^1 spectra. The CSP values were plotted against the number in the protein sequence (that the cross-peak relates too) to provide a visual illustration of the changes that are occurring in the protein, Figure 4.20. The CSP value is considered to be significant if it is greater than the average of the values plus 1 standard deviation [120]. Each CSP value is shown as a solid coloured bar in Figure 4.20 for CTPR3 (blue) and CTPRy3 (red). Where a cross-peak was present in the 0 M spectra but absent from the 1.6 M spectra there is an unshaded coloured bar with a value of 0.20. Where there was no cross-peak to assign in either spectra a small black bar with a value of 0.05 was added. From comparing the C-terminal section of the proteins (residues 106-120), it is clear that CTPRy3 losses more cross-peaks from the C-terminal helix than CTPR3, indicating that the C-terminal helix unfolds at 1.6 M GuHCl. In addition there are more cross-peaks that undergo a significant change in CTPRy3 overall than in CTPR3. This suggests that the CTPRy3 structure 'flexes' when the C-terminal helix is unfolded. When we combine the observation of the lost peaks with the CSP values we can conclude that the CTPRy3 unfolds its C-terminal helix.

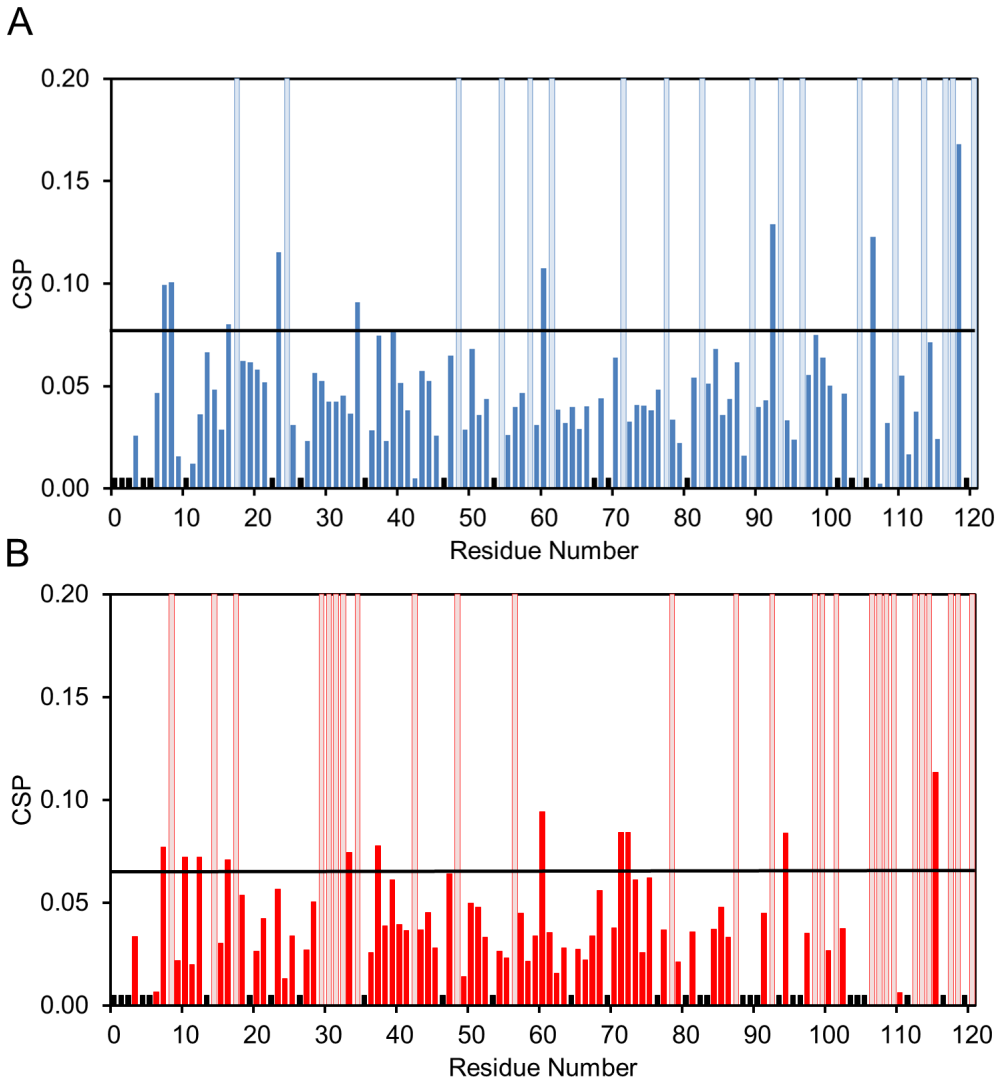

Figure 4.20: Chemical Shift Perturbation values for each cross-peak in CTPR3 (A) and CTPRy3 (B). The CSP values were calculated to quantitatively compare the change in chemical shift for each cross-peak that was observed between the HSQC spectra with 0 M and 1.6 M GuHCl. Each solid coloured bar is the CSP value calculated for each specific amide cross-peak. The CSP values are calculated by inputting the change (i.e. the difference between the 0 M and 1.6 M GuHCl) in chemical shifts for each cross-peak into Equation 4.8. Thus each CSP value represents a change in chemical shift for an assigned cross-peak. The unshaded coloured bars signify cross-peaks that are present in the 0 M spectra but absent from the 1.6 M spectra. The black bars are residues for which there is no cross-peak to assign in either spectra. The higher a CSP value is the greater the change in chemical shift. A significant change in CSP values is defined as a value higher than the average of the shifts plus 1 standard deviation. The horizontal black line marks this point. In the data for CTPR3 the value is 0.078 and in CTPRy3 it is 0.065. For CTPR3 we observe that there are less cross-peaks that have a significant shift than there are in CTPRy3. In addition there are many peaks that are lost from the CTPRy3 spectra, particularly those relating to the C-terminal helix (residues 106-120).

4.3.8.5 Conclusions

The CD wavelength scans show that at 0 M GuHCl, the molar ellipticity of CTPRy3 equals that of CTPR3, however at 2 M GuHCl the ellipticity of CTPRy3 is the same as CTPR2ΔS. This indicates that CTPRy3 has partially unfolded to give a structure that is identical to CTPR3ΔS. The chemical denaturation data for the CTPRy3 was compared with the simulated data from the Heteropolymer Ising model and was found to be almost indistinguishable, this confirms that this new version of the Ising model can accurately describe the denaturation curves of CTPRs. When comparing 2D NMR data, CTPRy3 showed specific unfolding of the C-terminal helix at 1.6 M GuHCl. From comparing both equilibrium unfolding data and 2D NMR HSQC experiments of CTPR3 and CTPRy3 we can see that CTPRy3 does possess a thermodynamic switch.

4.3.9 CTPRy2

Following the success of the switch design of CTPRy3 it is interesting to compare the properties of the same mutation in CTPR2. Figure 4.21 shows the location of the NN→RS mutation in the crystal structure of CTPR2.

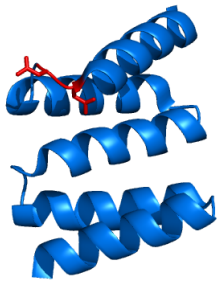

Figure 4.21: Crystal structure of CTPR2 (1NA3). Highlighted in red are the two asparagine residues that are to be mutated to arginine and serine to create the mutant CTPRy2.

4.3.9.1 Cloning, Expression and Purification of CTPRy2

CTPRy2 was cloned as described in Section 2.1 and Table 2.1, the mutations were confirmed via DNA sequencing. The protein was expressed and purified under native conditions as described in Section 2.3.2 and gave high yields of approximately 10 mg/L of culture.

4.3.9.2 Characterisation of CTPRy2

CTPRy2 was analysed using size exclusion chromatography to confirm that it is monomeric. The chromatogram for CTPRy2 was compared with that of CTPR2 (Figure 4.22 A. The mutant was confirmed as monomeric. A 100 μl sample at a concentration of 50 μM was loaded onto a GE Healthcare Superdex 75pg 300/10 column (equilibrated with 50 mM phosphate pH 7, 150 mM NaCl). The flow rate was set at 0.5 ml/minute. The elution of the protein was monitored using UV ab-

sorbance at 280 nm.

Far-UV CD wavelength scans of CTPRy2 were performed to assess the secondary structure, Figure 4.22 B. The wavelength scan was performed on a 5 μM sample, dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. The scans showed a single negative peak at 222 nm, therefore CTPRy2 displays the highly α-helical fold that is typical of CTPR proteins. The values for the molar ellipticity (expressed as mdeg $cm^{-2}dmol^{-1}$) were CTPR -1.72x10^5 and CTPRy2 -1.63x10^5. As the values are so close the CTPRy2 construct has the same helical content as CTPR2 and thus is correctly folded.

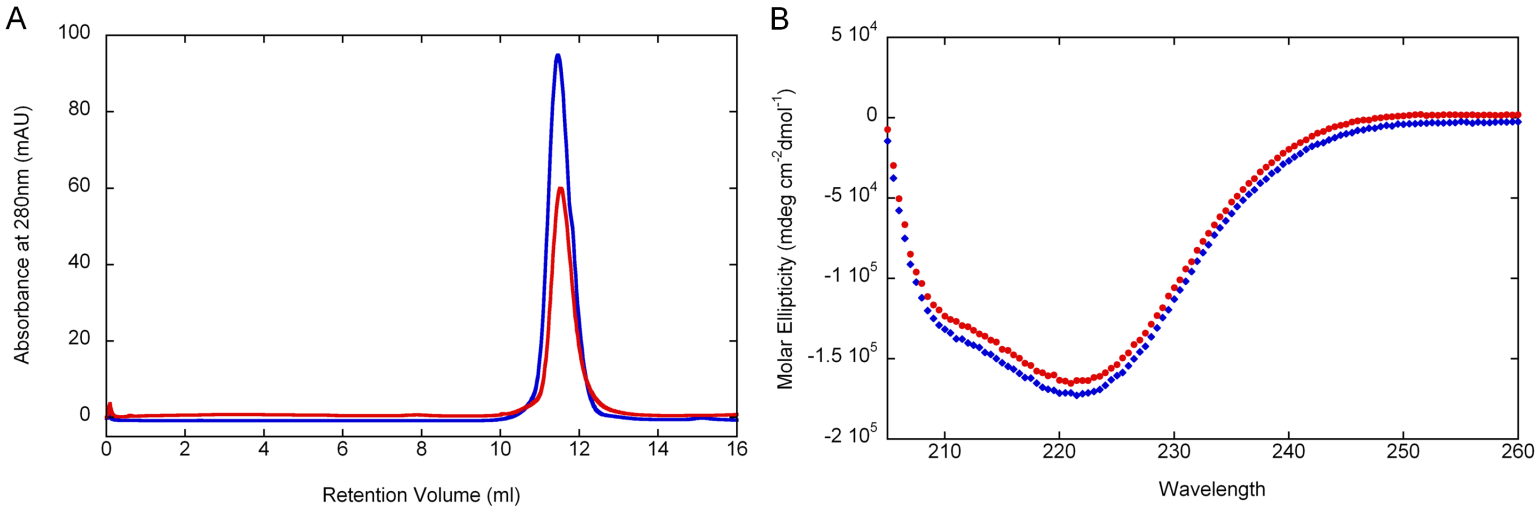

Figure 4.22: (A) Analytical size exclusion chromatograms of CTPR2 (blue) and CTPRy2 (red). Conditions were: 50 mM phosphate pH 7, 150 mM NaCl, buffer flow rate was set at 0.5 ml/minute. The column used was a GE Healthcare Superdex 75pg 300/10.The elution of the protein was monitored using the absorbance at 280 nm and the retention on the column measured in ml. The retention time is the same for both proteins, this shows that CTPRy2 is monomeric. (B) Far UV wavelength scan of CTPR2 (blue) and CTPRy2 (red) at 5 μM dissolved in 50 mM phosphate pH 7. The sample was held in a thermostated cuvette holder at 10 °C. The scan of CTPRy2 showed a single negative peak at 222 nm, indicating that the protein is highly α-helical.

4.3.9.3 Chemical Denaturation of CTPRy2

The stability of CTPRy2 was assessed in the same way as CTPRy3, see Section 4.3.8.3. CTPRy2 underwent a single transition from the folded to the unfolded state, Figure 4.23. The data for CTPRy2 can be qualitatively compared to data for the other CTPR2 mutants by fitting to a two-state two-sloping baselines equation 4.2.

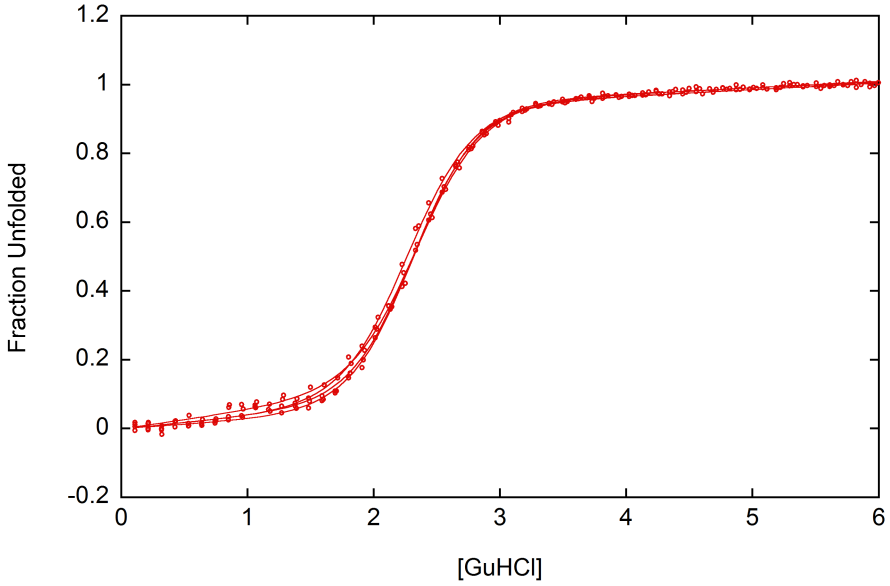

Figure 4.23: Normalised chemical denaturation curves for CTPRy2 (red circles), there are 4 data sets. The protein undergoes a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baselines equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation that arises in concentration between data sets. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 ^{o}C.

4.3.9.3.1 Comparison of CTPRy2 with CTPR2 and CTPR2ΔS

Figure 4.24A shows the normalised chemical denaturation curves for CTPRy2, CTPR2ΔS and CTPR2. The data for CTPRy2 can be qualitatively compared to data for the other CTPR2 mutants by fitting to a two-state two-sloping baselines equation 4.2. The mid-points of the fit of the data can be used to compare the stability of the different proteins. As with all other constructs the denaturations show one major reversible unfolding transition with increasing GuHCl. As predicted for CTPRy2, this transition has a lower midpoint than CTPR2, showing it is less stable. Moreover, there is an observable change in the slope of CTPRy2's folded baseline in comparison to CTPR2, Figure 4.24B. CTPR2's baseline is essentially flat, whereas CTPRy2's slopes towards the first transition. In addition, the midpoint of CTPRy2 is the same as that seen for CTPR2ΔS. This indicates that the CTPRy2 unfolds its C-terminal helix before the first the main transition, as it was designed to do.

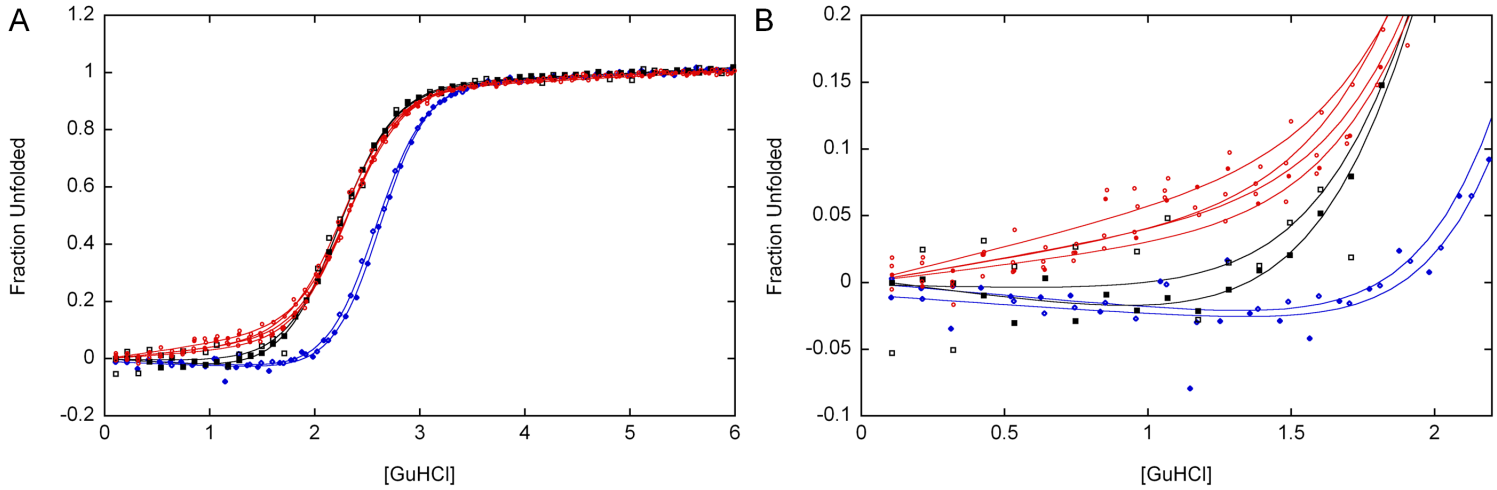

Figure 4.24: (A) Normalised chemical denaturation curves for CTPR2 (red circles), CTPR2 (Blue diamonds) and CTPR2ΔS (black squares). All proteins undergo a single reversible transition from the folded to unfolded state. Initially, the data was fit using a two-state two-sloping baselines equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation in ellipticity that arises from the different number of helices for each construct. Data was fit using linear regression analysis in Kaleidagraph 4.0. (B) Shows an enlargement of the folded baseline, showing that the baseline of CTPRy2 has a much steeper gradient than either CTPR2 or CTPR2ΔS. The data shows that the stability increases with the size of the protein, as judged by the mid-point of the denaturation. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 °C.

4.3.9.3.2 Chemical Denaturation of CTPRy2 in MOPS

Denaturation experiments performed in MOPS buffer (instead of phosphate) all showed a shift in the mid-point to a lower concentration of GuHCl. To confirm that this trends also applies to CTPRy2 the denaturation was repeated in 50 mM MOPS pH 7, Figure 4.25. As with all the other proteins examined in this chapter, there was a shift in the mid-point to a lower concentration of GuHCl by approximately 0.15 M GuHCl.

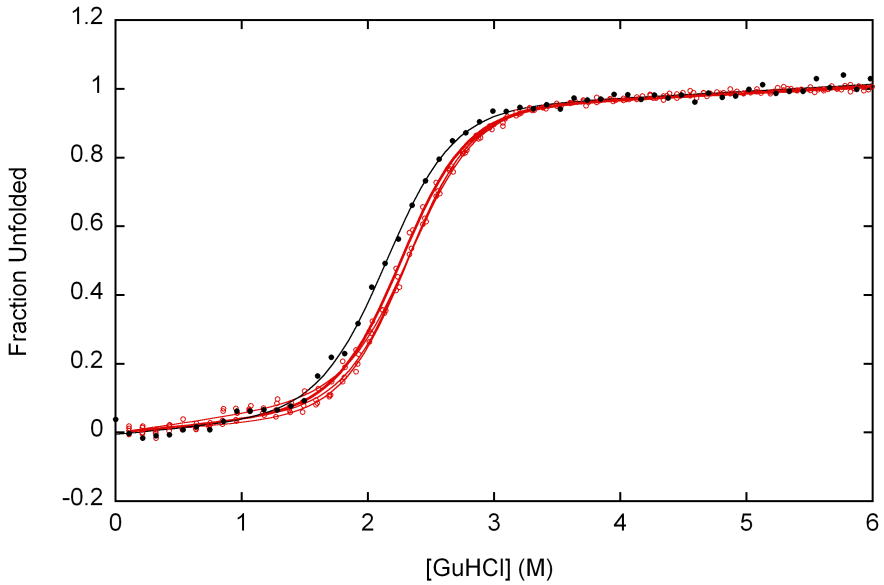

Figure 4.25: Normalised chemical denaturation curves for CTPRy2, in 50 mM MOPS pH 7 (black) and in 50 mM phosphate pH 7 (red). All data sets show that the protein undergoes a single reversible transition from the folded to unfolded state. The MOPS denaturation shows the same transition, but has a lower mid-point. This shift of approximately 0.15 M GuHCl is in agreement with the changes observed with the other CTPR proteins. Initially, the data was fit using a two-state two-sloping baselines equation (Section 2.9). However, to allow for easier comparison the data has been normalised (using Equation 4.1) to account for the variation that arises in concentration between data sets. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in either 50 mM MOPS pH 7 or 50 mM phosphate pH 7. GuHCl was used to denature the protein and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 ^{o}C.

4.3.9.4 2D NMR of CTPRy2 and CTPR2

N^{15} labelled CTPRy2 and CTPR2 were expressed and purified as described in Section 2.3.7.2. Samples were prepared for analysis in the same way as for CTPRy3, see Section 4.3.8.4 for details. Figure 4.26 shows an overlay of the spectra for both proteins.

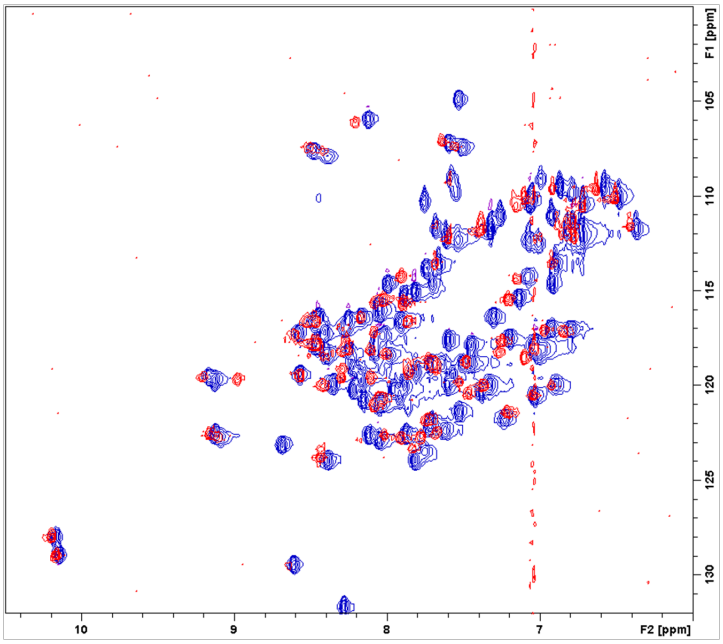

Figure 4.26: 2D HSQC NMR spectra of CTPR2 (blue) and CTPRy2 (red). Buffer conditions were 50 mM phosphate pH 6.8, 150 mM NaCl, with 10% (v/v) of D2O.

4.3.9.4.1 Assignment of CTPRy2 and CTPR2 spectra

Spectra were processed using TopSpin 2.1 patch level 6. Cross-peaks for CTPR2 were assigned by comparing spectra collected for CTPR2 with published assigned data, 77 peaks were assigned out of 86 residues. CTPRy2 was assigned by comparing CTPRy2 with the CTPR2 data, 62 peaks were assigned out of 86 residues. Figure 4.26 shows an overlay of the two spectra for CTPR2 and CTPRy2 at 0 M GuHCl. There is very little difference between the two spectra confirming that CTPRy2 adopts the desired CTPR fold.

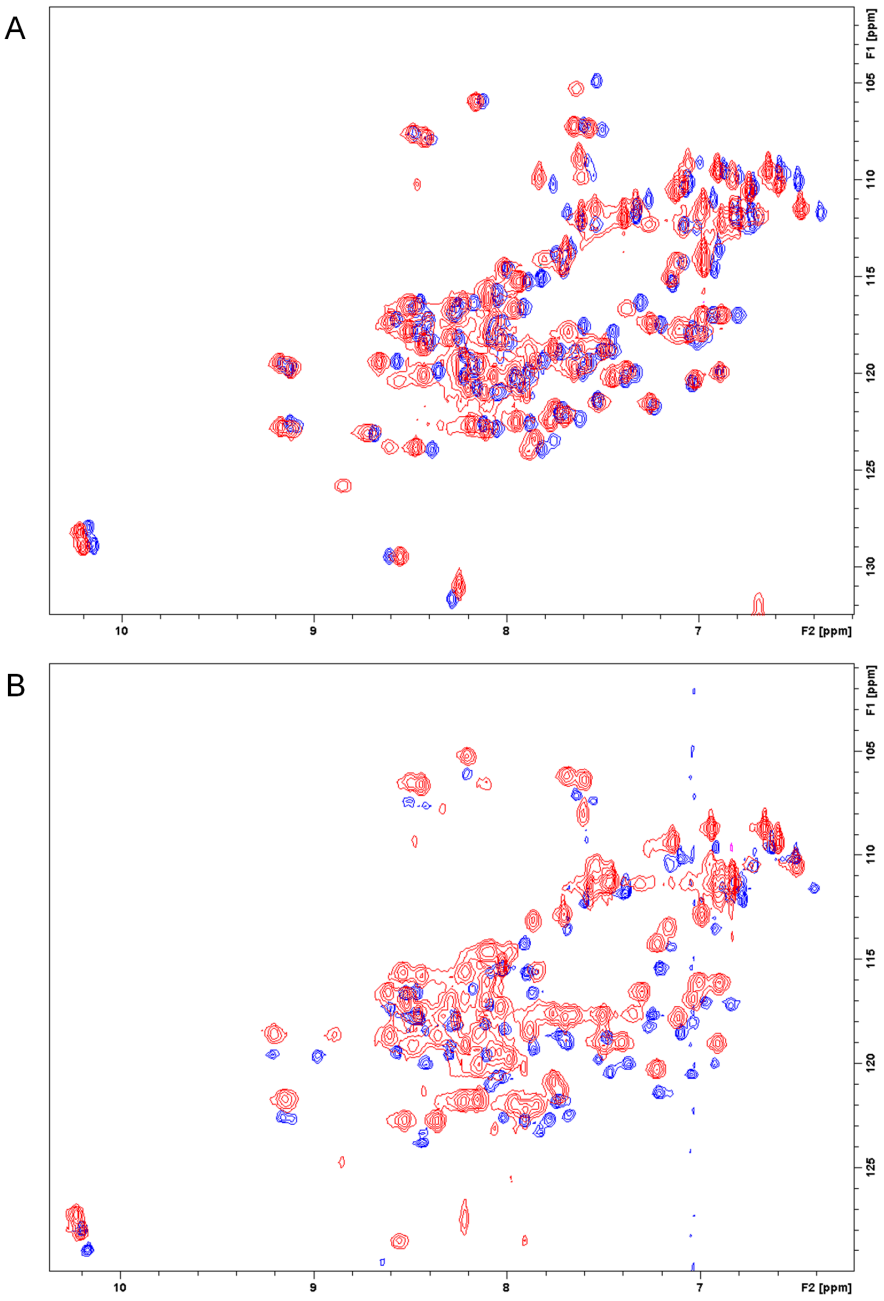

Figure 4.27: (A) 2D NMR HSQC data for CTPR2 in 0 M GuHCl (blue) and 0.8 M GuHCl (red). (B) 2D NMR HSQC data for CTPRy2 in 0 M GuHCl (blue) and 0.8 M GuHCl (red). Buffer conditions were the same for both samples: 50 mM phosphate pH 6.8, 150 mM NaCl. Data was recorded at room temperature.

The proteins were partially denatured with GuHCl and the 2D NMR HSQCs recorded again, Figure 4.27. The assignment of the partially denatured data (red) was accomplished through comparison with the 0 M sample (blue).

4.3.9.4.2 Comparison of CTPR2 and CTPRy2 2D NMR spectra

The HSQC spectra for both CTPR2 and CTPRy2 were well dispersed and showed little change between the two constructs. However, when the proteins were denatured with 0.8 M GuHCl, 6 % of cross-peaks were lost from the CTPR2 spectra (5 out of 77), compared with 10 % of cross-peaks (6 out of 62) were lost from the CTPRy2, Figure 4.28. The response of the C-terminal helix in CTPR2 showed that 13 % of peaks were lost (2 out of 15). The C-terminal helix of CTPRy2 was found to undergo greater structural changes, 29 % of cross-peaks were lost (2 out of 7 peaks).

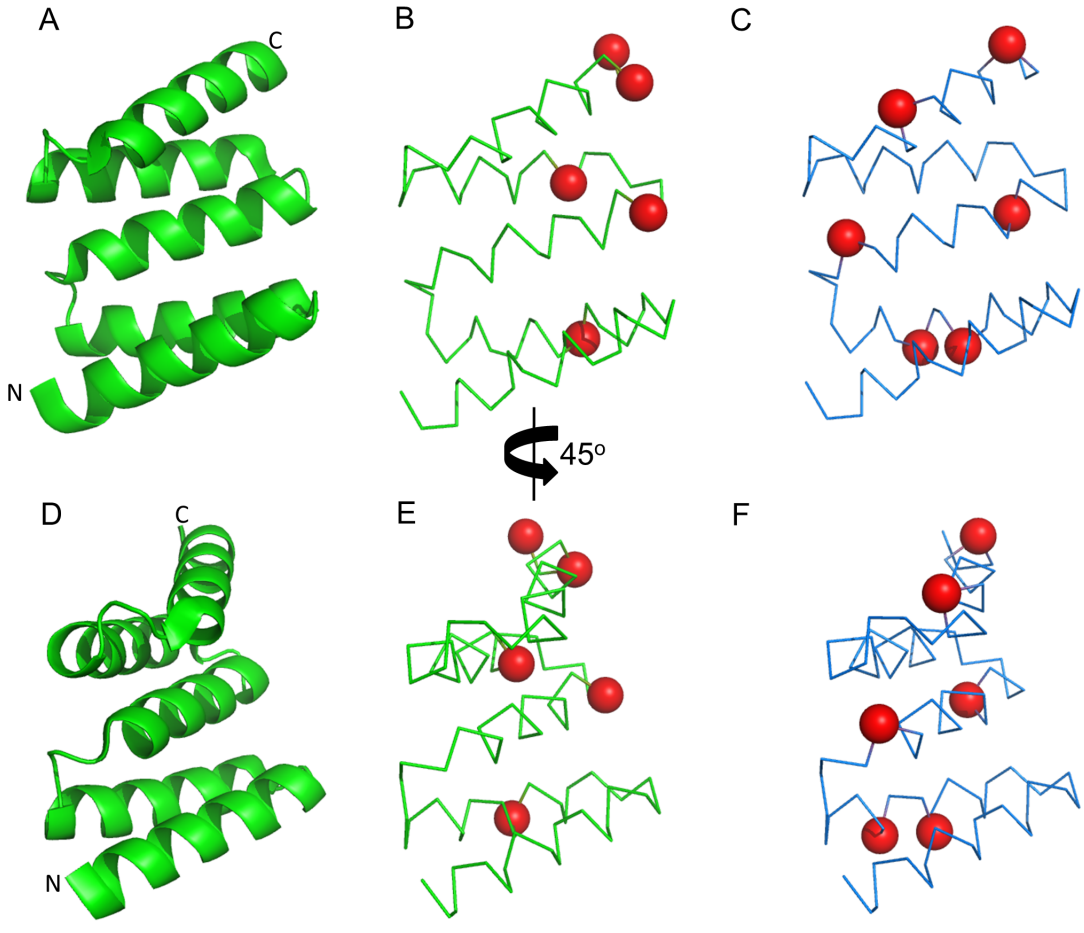

Figure 4.28: Changes in the HSQC spectra (between 0 M and 0.8 M GuHCl) mapped onto the structure of CTPR2 (A) and CTPRy2 (D) (1NA3) . Cross-peaks that are absent from the 0.8 M spectra are marked using a red sphere. CTPR2 (B and E) loses 5 peaks, where as CTPRy2 (C and F) loses 6 peaks.

Chemical shift perturbations (CSP) values were used to quantify the changes observed for individual cross-peaks between the two spectra. For each cross-peak, the change in chemical shift for both the N^{15} and the H^{1} spectra is calculated. These values are used in Equation 4.8 to calculate the CSP values. CSP values were calculated for each of the cross-peaks that were present in both the 0 M and 0.8 M GuHCl HSQC spectra, Figure 4.29. These values were plotted against the number in the protein sequence (that the cross-peak relates too) to provide a visual illustration of the changes that are occurring in the protein. The CSP value is considered to be significant if it is greater than the average of the values plus 1 standard deviation [120]. Each CSP value is shown as a solid coloured bar in Figure 4.29 for CTPR2 (blue) and CTPRy2 (red). Where a cross-peak was present in the 0 M spectra but absent from the 0.8 M spectra there is a unshaded coloured bar with a value of 0.20. Where there was no cross-peak to assign in the either spectra a small black bar with a value of 0.05 has been added. We can see that the CTPR2 shows only small changes in the CSP values when compared with CTPRy2, indicating that CTPR2 remains folded at 0.8 M GuHCl. The larger perturbations that are seen throughout CTPRy2 suggest that the whole protein structure begins to 'flex' in response to 0.8 M GuHCl.

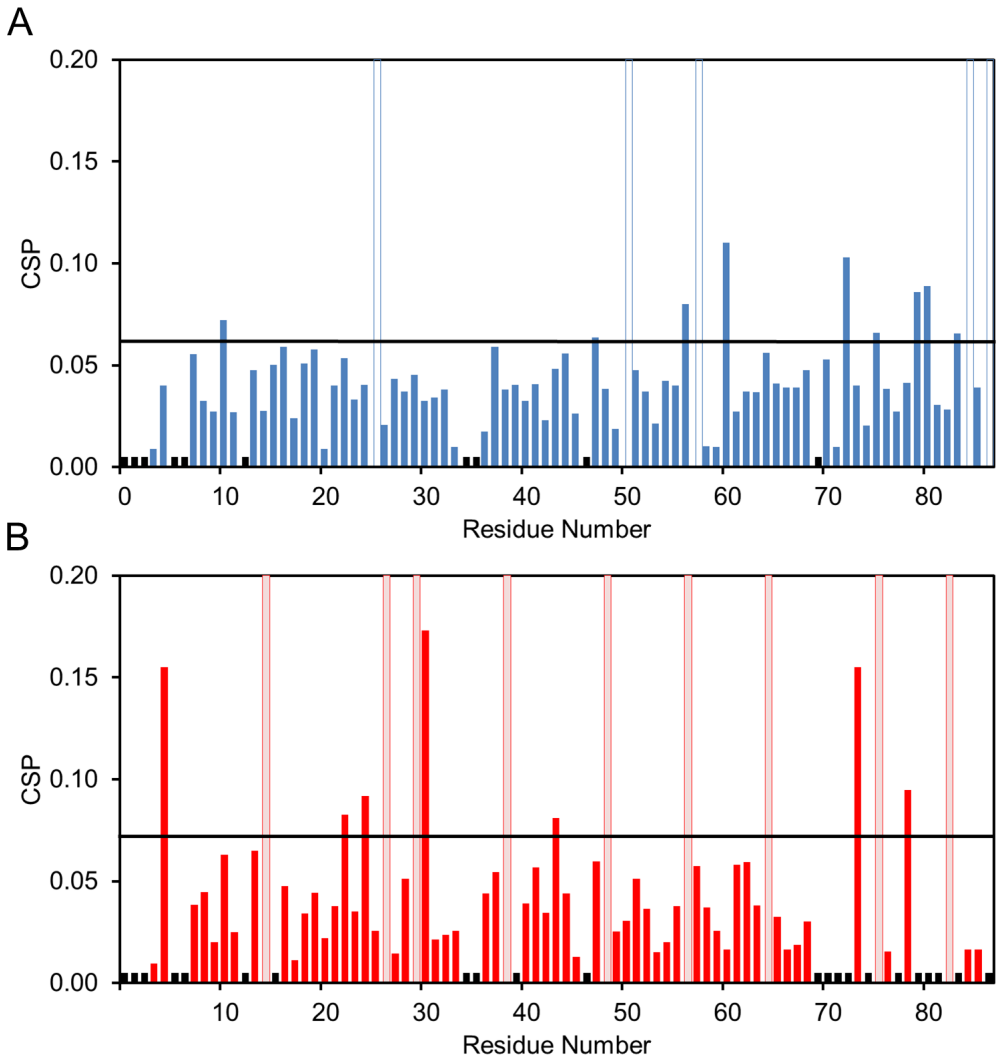

Figure 4.29: Chemical Shift Perturbation values for each cross-peak in CTPR2 (A) and CTPRy2 (B). The CSP values were calculated to quantitatively compare the change in chemical shift for each cross-peak, that was observed on addition of 0.8 M GuHCl. Each solid coloured bar is the CSP value calculated for each specific amide cross-peak. The CSP values are calculated by inputting the change (i.e. the difference between the 0 M and 0.8 M GuHCl) in chemical shifts for each cross-peak into Equation 4.8. Thus each CSP value represents a change in chemical shift for an assigned cross-peak. The unshaded coloured bars signify cross-peaks that are present in the 0 M spectra but absent from the 0.8 M spectra. The black bars are residues for which there is no cross-peak to assign in either spectra. The higher a CSP value is the greater the change in chemical shift. A significant change in CSP values is defined as a value higher than 'the average of the shifts plus 1 standard deviation'. The horizontal black line marks this point. In the data for CTPR2 the value is 0.062 and in CTPRy2 it is 0.073. We observe that there are almost the same number of cross-peaks that display a significant change. However there is a larger perturbation in general in the cross-peaks throughout the CTPRy2 protein, suggesting that the whole protein structure of CTPRy2 begins to 'flex' in response to 0.8 M GuHCl.

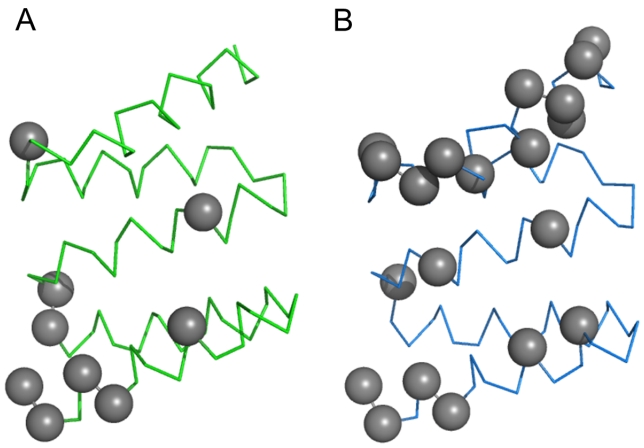

Figure 4.30: Residues for which there is not a corresponding cross-peak have been mapped onto the structure of CTPR2 (1NA3) as a grey sphere for CTPR2 (A) and CTPRy2 (B). There are 11 less peaks in the CTPRy2 spectra than there are present in CTPR2. The most noticeable difference is that there are 8 less peaks in the C-terminal helix in the spectra of CTPRy2 compared with CTPR2. This means that more than 50 % of the C-terminal helix is not visible in the NMR experiments.

The main limitation when observing the response of CTPRy2 to GuHCl is that there are so few cross-peaks for the C-terminal helix even at 0 M GuHCl. In comparison, CTPR2 has cross-peaks for all residues in the C-terminal helix. Figure 4.30 shows the residues that are not visible in the 2D NMR spectra at 0 M GuHCl. Here we can see that there are fewer probes available in the CTPRy2, especially in the C-terminal helix. The consequence of this is using 2D NMR to probe the structural response of CTPRy2 to denaturant is severely hindered.

4.3.9.5 Conclusions

When we look at the response of CTPRy2 using CD equilibrium unfolding we can see that there is a difference in the folded baseline compared with CTPR2 and CTPR2ΔS. This indicates that the C-terminal helix is unfolding. 2D NMR data shows that CTPRy2 displays the same structure as CTPR2. However 2D NMR (under the conditions used) was not suitable to probe the structural changes that CTPRy2 undergoes.

4.3.10 Comparison of the Switching Ability of CTPRy3 and CTPRy2

When we compare the response of the CTPRy2 to the CTPRy3 we see that there is a much greater response from the CTPRy3. We are able to achieve an 83 % loss of structure in the solvating helix whilst maintaining a core structure that only sees a loss of 23 %. However, with the CTPRy2 the response at low concentrations indicates some switching ability, increasing the denaturant concentration doesn't increase the loss of the solvating helix but does increase the loss of cross-peaks associated with the core of the protein. We can conclude that the core of the protein structure begins to 'flex' following the significant loss of structure due to the unfolding of the C-terminal helix.

4.4 Conclusions

In this chapter I have described how a series of deletion mutants was used with a Heteropolymer Ising model to design a 'switching' system. The newly designed protein, CTPRy3, reversibly populates a partially folded stable intermediate. The thermodynamic switch was designed to selectively unfold the C-terminal S helix (S-switch). This was chosen as the target for 3 reasons: (a) we have previously shown that proteins that lack the C-terminal helix are stable (Chapter 3); (b) we already had a potential way of destabilising the C-terminal helix, as from previous work we know that the `NN`→`RS` mutation is destabilising and (c) removal of this helix reveals a potential dimerisation interface. This is important if we want to induce oligomerisation following partially unfolding the CTPR, as for this to occur a potential stacking interface needs to be exposed.

Whilst there were several possible strategies for unfolding the S helix, it seemed logical to start with a mutation that is known to reduce stability in the CTPR proteins. The construct CTPRy3 combines the more stable core of the CTPR series with the less stable inter-loop region of the CTPRa series, placed after the final repeat but before the C-terminal helix. As CTPRy3 is a combination of these two extensively studied protein series, we have been able to use the Heteropolymer Ising model to simulate the response to chemical denaturation. When modelled, the new CTPRy3 construct showed more potential as a thermodynamic switch than either CTPR3 or CTPRa3.

To test the predictions made by the Heteropolymer Ising model, the proteins CTPRy3 and CTPRy2 were cloned and expressed. When these proteins were thermally and chemically denatured the data showed a reduction in overall stability compared with CTPR2 and CTPR3. Further analysis via 2D NMR has revealed that the

CTPRyn proteins can selectively unfold the S helix at a low GuHCl concentrations whilst the core of the protein remains folded. We have found the CTPRy3 protein to exhibit a more defined 'switch' than the CTPRy2 construct, it is likely that this is due to the added stability gained from CTPRy3 being 1 repeat longer than CTPRy2. As increasing the number of repeats increases stability in both CTPR/CTPRa series therefore having an extra repeat will give added stability to the partially folded intermediate.

Chapter 5

Photoisomerisable Azobenzene Crosslinker

In this chapter, I will describe the development of a light-sensitive protein switch. This conformational switch will use the photoisomerisable azobenzene linker; 3,3'-Bis(sulfo)-4,4'-bis chloroacetamidoazobenzene (referred to as AB-photolinker). This is bonded to a CTPR3 scaffold through two cysteines in the C-terminal helix. When exposed to UV light at 365 nm the isomerisation of the linker will unfold the C-terminal helix.

5.1 Azobenzene linker

Azobenzene molecules can undergo a reversible cis-trans isomerisation, see Figure 5.1. The trans→cis isomerisation is accomplished using a specific wavelength of light, that is dependent on any substituents on the phenyl rings and the solvent conditions. The reverse cis→trans isomerisation can be accomplished using either UV light or heat. The trans conformation is more thermodynamically stable so the cis→trans conversion will occur over time if the sample is left in the dark [121].

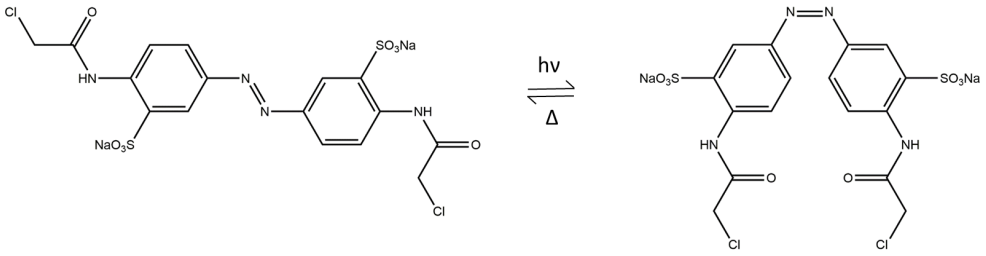

Figure 5.1: Shows the trans to cis isomerisation of the AB-photolinker (3,3'-Bis(sulfo)-4,4'-bis chloroacetamidoazobenzene), using UV light at 365 nm. The reverse reaction is achieved thermally.

When crosslinked through 2 side chains, azobenzenes have been used to cause a conformational change in biomolecules. One azobenzene linker was used to inhibit the activity of ion channels, by either blocking the channel or causing a conformational change that deactivated the channel [122]. The azobenzene linker used for this project was 3,3'-Bis(sulfo)-4,4'-bis chloroacetamidoazobenzene (AB-photolinker), Figure 5.1. It was chosen as it had been shown to disrupt the helical content of an α-helical peptide [123].

The AB-photolinker has been modified through the addition of the 2 sulfonate groups. As a result, the AB-photolinker is soluble in water whilst maintaining a similar size to the non-sulfonated linker. Interestingly, the addition of the sulfonate group was found to increase the half-life of the cis$\rightarrow$trans isomerisation [123]. This is beneficial as it increases the time available to detect any change in the protein structure. The AB-photolinker used in this experiment has a half-life of 35 minutes at room temperature.

5.2 Experimental Approach

Four CTPR3 mutants were designed to possess 2 cysteine mutations in the C-terminal helix. Sites were selected to allow attachment of the AB-photolinker. These mutants were cloned, expressed and purified. The AB-photolinker was ligated to

the CTPR3 scaffold via the cysteine residues. UV wavelength scans and Circular Dichroism scans were used to investigate if the isomerisation of the linker results in a conformational switch in the proteins.

5.3 Motivation for creating a photo-switchable CTPR

In Chapter 4 I described how a conformational switch was designed into CTPR2 and CTPR3. This switch used GuHCl to trigger the structural change. Here we aim to accomplish the same switch but under the control of UV light. Figure 5.2 shows the proposed mechanism for the conformational switch that is initiated by the AB-photolinker.

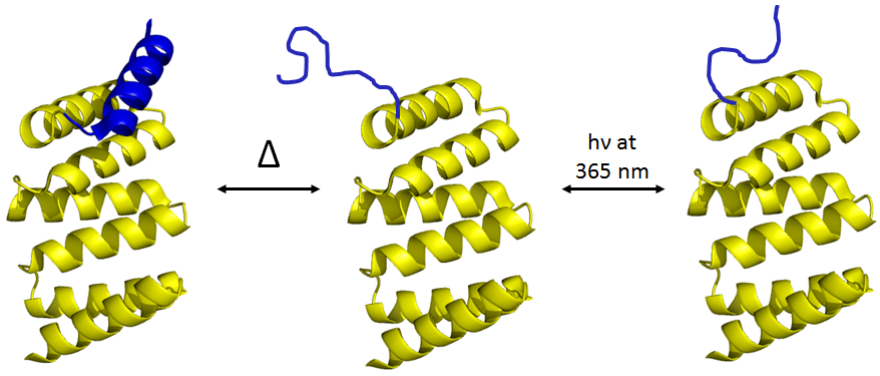

Figure 5.2: The AB-photolinker is attached through 2 cysteines on the C-terminal helix (blue) in the trans conformation. After partial denaturation using heat the C-terminal helix partially unfolds. UV light at a wavelength of 365 nm is used to isomerise the AB-photolinker into the cis conformation, keeping the C-terminal helix unfolded. Constructed from 1NA0.

5.4 Results

5.4.1 Selection of Mutation sites

The trans isomer of the AB-photolinker requires a spacing of 10-11 residues between the cysteines to which it will be linked. In addition to the spacing, mutation sites were selected so that both cysteines are solvent exposed. Four constructs were proposed (see Figure 5.3). The sequences of the C-terminal helices for each mutant are listed below. Two variations of each cysteine mutant were produced: one with the destabilising `PNN-PRS` mutation and one without. As we have shown in chapter 4, the `PNN-PRS` mutation destabilised the C-terminal helix relative to the rest of the protein and by adding it here we aimed to increase the percentage of the population that undergoes a conformational change.

A

CTPR3	P	N	N	A	E	A	K	Q	N	L	G	N	A	K	Q	K	Q	G
CTPR3i4	P	N	N	A	E	A	K	C	N	L	G	N	A	K	Q	K	Q	C
CTPRy3i4	P	R	S	A	E	A	K	C	N	L	G	N	A	K	Q	K	Q	C
CTPR3i11	P	N	N	C	E	A	K	Q	N	L	G	N	A	K	C	K	Q	G
CTPRy3i11	P	R	S	C	E	A	K	Q	N	L	G	N	A	K	C	K	Q	G

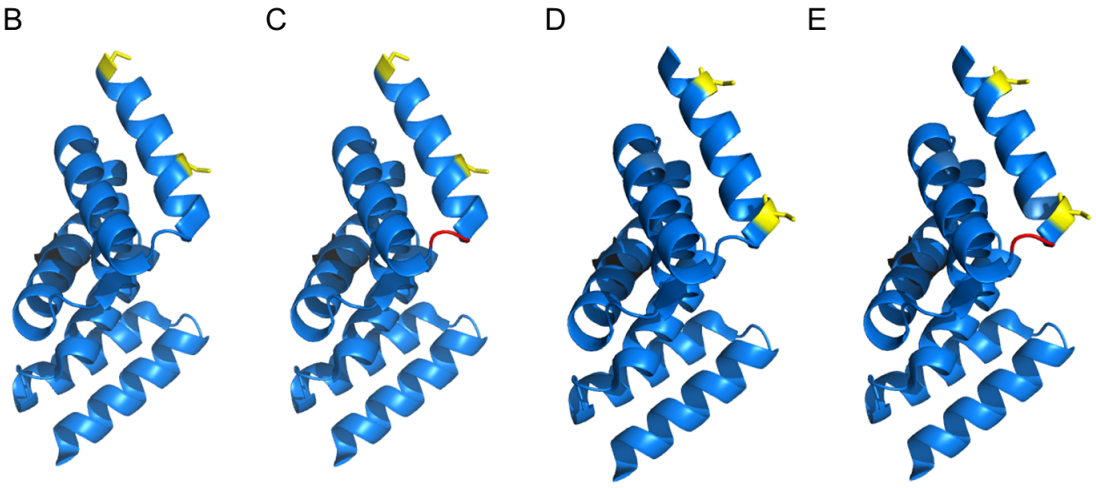

Figure 5.3: (A) Four double cysteine mutants were designed, the sequence of the C-terminal helix is shown with the WT CTPR3 sequence. Structure of CTPR3 (1NA0) with the mutations highlighted for each construct: (B)CTPR3i4 (C)CTPRy3i4 (D)CTPR3i11 (E)CTPRy3i11. The residues mutated to cysteine are highlighted in yellow. The PRS mutation is highlighted in red.

5.4.2 Cloning, Expression and Characterisation of Cysteine Mutants

The four double cysteine mutants were cloned as described in Section 2.1 and Table 2.1. The mutations were achieved through Site-directed Mutagenesis as described in Section 2.2.2. The mutations were confirmed via DNA sequencing. Proteins were expressed and purified under native reducing conditions as described in Chapter 2.3.2. All give high yields of approximately 10 mg/L of culture. All proteins were monomeric as judged by the chromatograms for the size exclusion purification (data not shown).

5.4.2.1 Denaturation of CTPR3i11 and CTPR3yi11

To assess the effect of the cysteine mutations on the stability of the protein a chemical denaturation was performed as described in Section 2.5.2.2. The data was fit to a two-state model as described in Section 2.9, Figure 5.4. When compared with CTPR3 the cysteine mutant (CTPR3i11) is less stable as it has a lower mid-point. However it is interesting to observe that the CTPRy3i11 protein is more stable than the CTPRy3 construct, Table 5.1. As the constructs have similar stabilities to the CTPR3 and CTPRy3 scaffold we can be sure that conformational changes observed as a consequence of the AB-photolinker are not due structural changes as a result of the cysteine mutations.

Table 5.1: Comparison of Chemical denaturation data

Construct	**Chemical Denaturation Data**		
	[a] $[D]_{50\%}$ (M)	[a] m-value $kcal^{-1}M^{-1}$	[b] $\Delta G^{H_2O}_{D-N}$ $kcal^{-1}$
CTPR3	3.29±0.01	3.61±0.13	11.9±0.4
CTPRy3	3.17±0.01	2.65±0.07	8.4±0.2
CTPR3i11	3.30±0.01	2.64±0.07	8.7±0.2
CTPRy3i11	3.25±0.01	2.96±0.08	9.6±0.3

Errors are calculated from the fit of the graph.
[a] Calculated from Equation 4.2
[b] Calculated from Equation 4.3

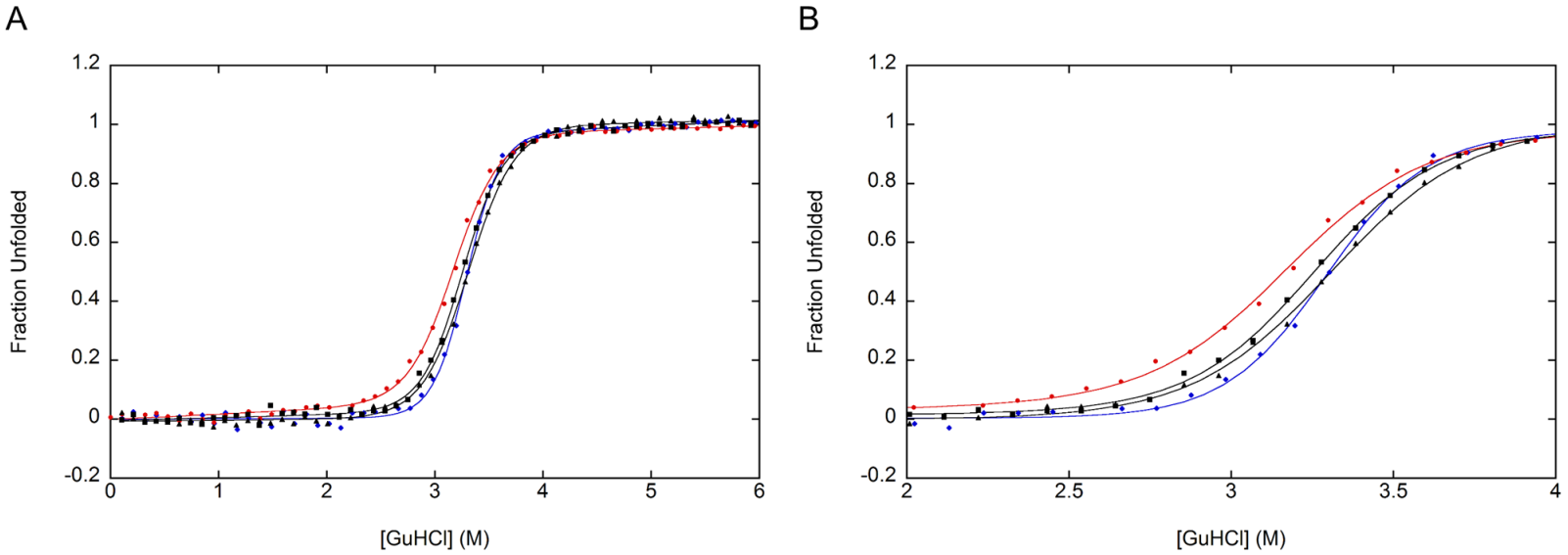

Figure 5.4: Chemical denaturations of CTPRy3 (red circles), CTPR3 (blue diamonds), CTPR3i11 (triangles) and CTPRy3i11 (squares). (A) is the complete denaturation curve and (B) is a enlargement of the transition. The cysteine mutations to appear destabilise the protein relative to the CTPR3. However they also have a paradoxical effect with respect to the PRS mutation, i.e. when the cysteine mutants are combined with the PRS mutation the resultant protein is more stable than the CTPRy3 protein that has only has the PRS mutation. All proteins undergo a single reversible transition from the folded to unfolded state. Data was fit using linear regression analysis in Kaleidagraph 4.0. Experimental conditions were: 5 μM protein in 50 mM phosphate pH 7. Guanidine HCl was used to denature the proteins and the progress was monitored using the CD ellipticity at 222 nm. Data was recorded on a Photophysics Chirascan using a 5 mm pathlength cuvette held in a thermostated holder set at 10 ^{o}C

5.4.3 Ligation of the AB-photolinker onto the CTPR3 scaffold

The protein was dialysed into 50 mM Tris pH 8, 2 mM TCEP. A concentrated stock solution of 5 mM AB-photolinker in ddH_2O was prepared immediately before use. In total, 500 nmol of linker was reacted with 90 nmol of protein (approximately a 5 fold excess of linker). The linker was added to the 900 μl of protein (100 μM) in 3 aliquots (30, 30 and 40 μl) with an incubation of 20 minutes between each addition. The reaction was performed in the dark, at room temperature, with shaking, this was to ensure the linker was in the trans conformation. The linker was added in 3 aliquots, to reduce the likelihood of intermolecular cross-linking.

5.4.3.1 Determining labelling efficiency using Ellman's reagent test for free thiols

The efficiency and yield of AB-photolinker attachment was assayed using Ellman's reagent. The test involves the use of the thiol reactive compound 5,5'-dithiobis-2-nitrobenzoic acid (DTNB). If free thiols exist in the protein, the DTNB reacts in a 1:1 ratio, releasing 1 mol equivalent of yellow 3-thio-6-nitrobenzoate (TNB), Figure 5.5. The TNB absorbs at 412 nm, therefore the number of free thiols can be calculated by using the Molar Extinction Co-efficient (14 150 $M^{-1}cm^{-1}$) of TNB and the Beer-Lambert law. When proteins CTPR3i4, CTPRy3i11 and CTPRy3i11 were crosslinked with the AB-photolinker, the Ellman's test showed that >80 % of the cysteines had reacted compared to the unlinked protein.

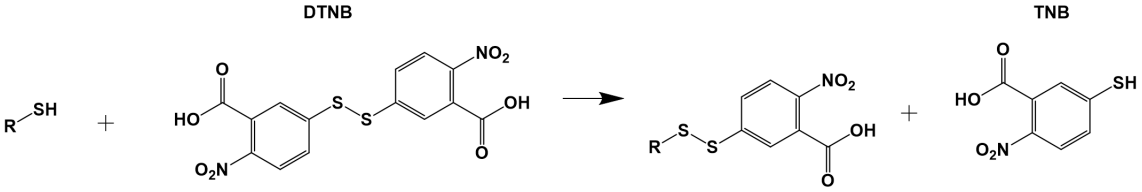

Figure 5.5: Reaction scheme of DTNB with a free thiol. The reaction produces 1 mol equivalent of the yellow TNB ion.

5.4.3.1.1 Mass Spectrometry to confirm linking

The linking reaction was confirmed using MALDI-TOF mass spectrometry. The samples were desalted using a Millipore C18 Zip-tip. The samples were mixed with a saturated Sinapinic acid solution in a 1:1 ratio before a 3 μl aliquot was spotted onto the MALDI plate. MALDI mass spectrometry of the AB-photolinked CTPRy3i11, CTPR3i4 and CTPR3i11 confirmed that the linker was attached (data not shown). However, unlinked protein was also observed on the spectra.

5.4.4 Isomerisation of AB-photolinked CTPR3

5.4.4.1 Isomerisation protocol

The isomerisation was accomplished using a 6 Watt 365 nm UV lamp. Initially the sample was heated at 50-60 ^{o}C for 5-10 minutes before isomerisation, as we believed the linker could not drive thermodynamic helix unfolding. However, this was found to be unnecessary and isomerisation could be achieved simply by exposure to UV light at 365 nm for 10-15 minutes.

5.4.4.2 UV Absorption characterisation

5.4.4.2.1 Methodology

UV absorbance wavelength scans were performed on a U-3010 Spectrophotometer scanning between 200-600 nm at room temperature. The protein concentration was between 15-25 μM in 50 mM Tris pH 8. Samples were measured in a 10 mm path-length cuvette. The trans isomer absorbs at 350 nm, thus the isomerisation reaction was monitored by observing the change in absorbance at this wavelength.

5.4.4.2.2 Results and Discussion

The UV spectra of CTPR3i4, CTPR3i11 and CTPRy3i11 all showed a decrease in the absorbance at 365 nm after exposure to UV light, Figure 5.6A, B and C. This indicates that the linker was isomerising from trans to cis. Figure 5.6D is the UV wavelength scan of the free AB-photolinker, showing how the linker behaves when not ligated to a protein. The absorbance at 365 nm significantly decreases by 70 %. All the AB-photolinker proteins show a percentage decrease of approximately 50 % Published data in reference [123], shows a higher decrease of approximately 70 %. However, in this example the linker is attached to a single helix.

5.4.4.3 Circular Dichroism characterisation

The UV absorbance spectra shows that the AB-photolinked protein is undergoing a trans→cis isomerisation. In Chapter 4 we used Circular Dichroism (CD) to identify changes in the structure of CTPRs, specifically the unfolding of the C-terminal helix at a specific concentration of GuHCl. In this chapter we are still investigating the unfolding of the C-terminal helix, albeit by the AB-photolinker rather than GuHCl, thus we consider CD a suitable detection method. The half-life for the relaxation of the AB-photolinker back to the trans conformation is 35 minutes at

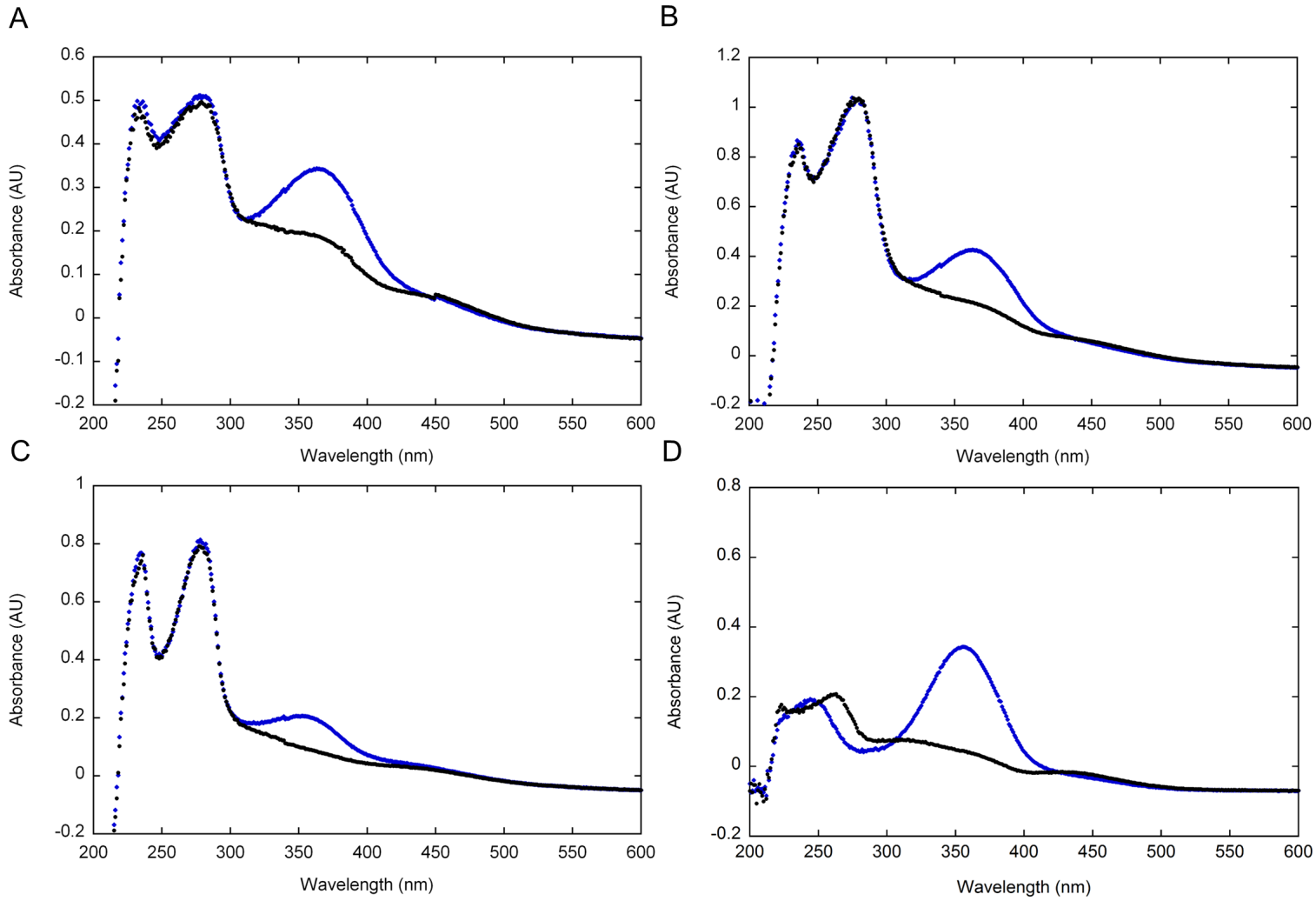

Figure 5.6: UV wavelength scans of CTPR3i4 (A), CTPR3i11 (B) and CTPRy3ill (C) before isomerisation, in the trans conformation (blue diamonds) and after isomerisation, in the cis conformation (black circles). All show approximately a 50 % decrease in the absorbance signal at 365 nm. The free linker (D) shows a much greater decrease in the signal after isomerisation.

room temperature. As the time taken to perform a CD wavelength scan is less than 5 minutes, the relaxation of the linker should have a negligible effect on the CD signal.

5.4.4.3.1 Methodology

Samples were diluted to a concentration of either 5 μM or 10 μM. The sample buffer was 50 mM Tris at pH 7 or pH 8. A 5 mm pathlength cuvette was used in a thermostated cuvette holder held at 10 °C. All CD scans were performed between 200-300 nm. An initial scan was performed in order to measure helical content of the fully folded protein. The sample was then isomerised using UV light at 365 nm as described in the previous section and another wavelength scan performed. The α-Helical content of the protein was analysed by observing the Ellipticity at 222 nm.

5.4.4.3.2 Results and Discussion

The isomerisation was conducted for the proteins: CTPR3i4, CTPR3i11 and CTPRy3i11. Figure 5.7 shows two indicative spectra for CTPRy3i11, which are representative of the other data sets. When the isomeriation was performed at pH 7 the CTPRy3i11 protein showed a reduction of 7.5 % of the CD signal. When the experiment was repeated at pH 8 a reduction of 4.2 % was recorded. The ideal theoretical reduction in CD signal would be 14 % but this would assume that 100 % of the protein population fully unfolds the C-terminal helix. These data show that there is a reduction in the CD signal at 222 nm which is indicative of a loss of helical structure as a result of the isomerisation of the AB-photolinker.

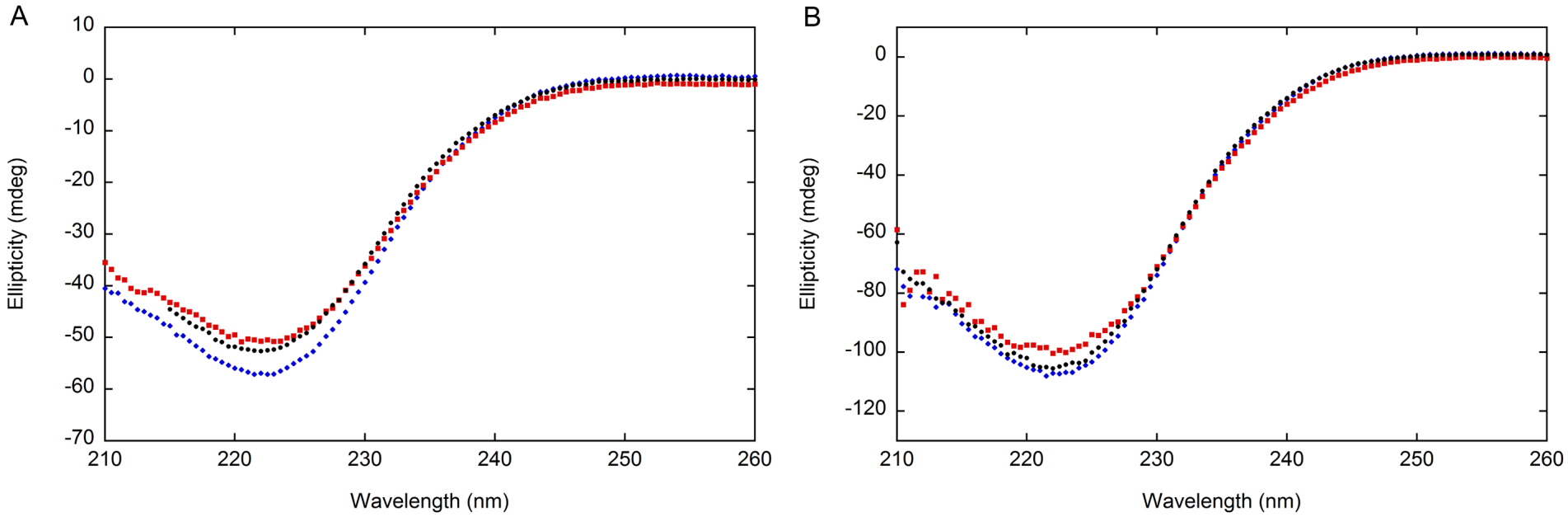

Figure 5.7: Far UV CD wavescans from the isomerisation of AB-photolinked protein CTPRy3i11 before isomerisation (blue diamonds), after heating (red squares) and after exposure to UV (black circles). (A) the sample has a concentration of 5 μM and is dissolved in 50 mM Tris pH 7. Here we can see prior to heating and UV activation the CD signal at 222 nm is -56.9 mdeg. After heating and UV activation for 10 minutes the signal has reduced to -52.6 mdeg, a 7.5 % reduction in signal. The signal change is indicative of a reduction in helical structure as a result of the isomerisation of the AB-photolinker. (B) The experiment was repeated using the same protein, CTPRy3i11, but at a concentration of 10 μM and dissolved in 50 mM Tris pH 8. Here the difference between the unisomerised and the isomerised is less pronounced (-110.6 mdeg and -105.9 mdeg respectively) with a reduction of 4.2 % of the CD signal observed after isomerisation.

5.4.4.4 2D NMR

2D NMR was used to further investigate if there are any structural changes occurring in the AB-photolinked protein. This can give a picture of any structural changes on a residue by residue basis. This was performed in a similar manner as described in Chapter 4 Section 4.3.8.4.

5.4.4.4.1 Methodology

N^{15} labelled CTPRy3i11 was prepared as described in Section 2.3.7.2. The AB-photolinker was ligated to the protein as described in Section 5.4.3. The photolinked protein was dialysed into 50 mM phosphate pH 6.8, 150 mM NaCl. The protein was concentrated to 500 μM. The isomerisation of the AB-photolinker was performed as described in Section 5.4.4. The isomerisation of the linker was confirmed by a decrease in the absorbance at 365 nm. The isomerised sample was kept at 4 ^{o}C with illumination from a UV lamp, until the NMR data was recorded. The sample was left at room temperature in the dark for approximately 5 hours, in order for the linker to thermally relax back to the trans configuration.

5.4.4.4.2 Results and Discussion

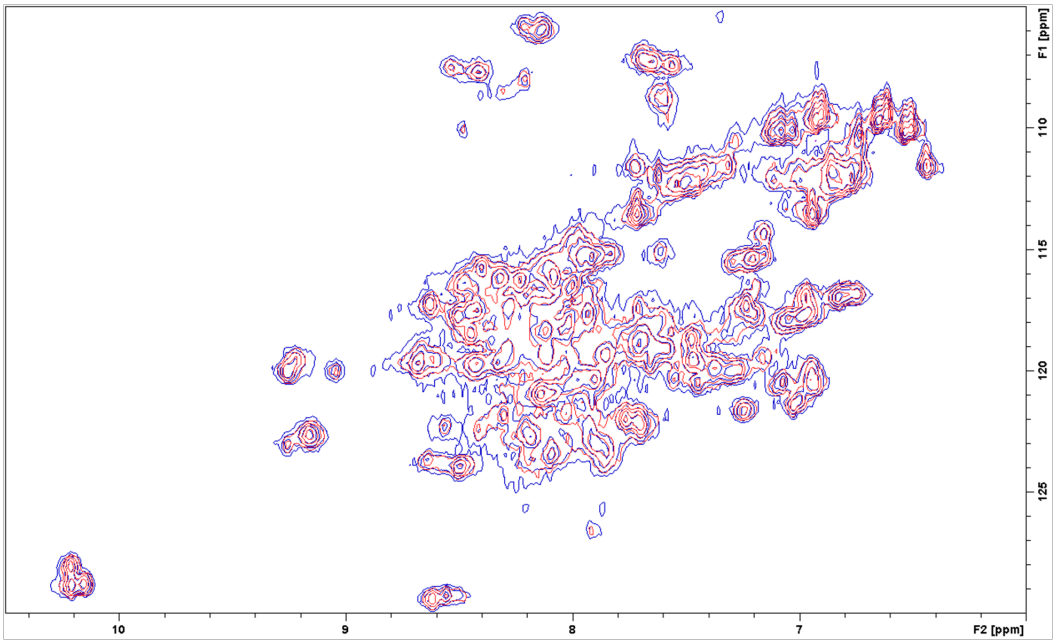

Figure 5.8: 2D HSQC for CTPRy3i11 before (blue) and after isomerisation (red). Buffer conditions were 50 mM phosphate pH 6.8, 150 mM NaCl, with 10% (v/v) of D2O.

The NMR data showed no changes in the spectra between the cis and the trans configuration of the AB-photolinker. Although a change in the UV spectra was observed (0.263 before UV exposure and 0.171 after exposure) there was no change in the NMR spectra. There were two potential reasons why no significant change in the either the CD ellipticity or NMR data was observed: (1) the linker is only isomerising a small number of the protein population or (2) the linker is isomerising, but is not causing a large enough change in the helical structure of the C-terminal helix to be observable. It seems likely that the isomerisation of the linker results in some disruption of the terminal helix but that there is not enough of a change to be visible either through CD or NMR.

To determine if the lack of change in the CD signal was the result of no structural changes in the proteins a new method of detection was devised. This involved producing a mutant of CTPRy3i11 with a cysteine in the last C-terminal repeat at

position I97. This was intended to show that the new cysteine residue is hidden in the core of the protein when the C-terminal helix is folded. However, when the C-terminal helix is unfolded the cysteine is exposed. This reaction was followed using Ellman's reagent.

5.4.5 Cysteine mutant CTPRy3i11 I97C

This mutant was designed so that the Ellman's Reagent could be used to identify if the C-terminal helix was unfolding in response to the trans to cis isomerisation of the AB-photolinker. The I97 residue was selected for mutation as it was calculated to have a large change in Solvent Accessible Surface Area (SASA) upon changing from CTPR3 to CTPR3ΔS. It was predicted that the cysteine residue would be inaccessible when the C-terminal helix is folded and exposed when the helix is unfolded. Upon exposure the cysteine will be free to undergo reaction with the Ellman's Reagent, which can be followed by measuring the absorbance at 412 nm.

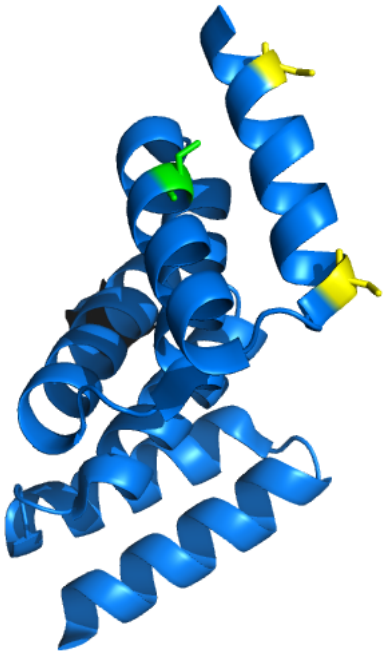

Figure 5.9: The CTPRy3i11 I97C protein. The double cysteine mutations that are for the ligation of the AB-linker are highlighted in yellow. The I97C mutation is highlighted in green. Constructed from 1NA0.

5.4.5.0.3 Methodology

To avoid wasting the AB-photolinker, which is not commercially available, maleimide was used to simulate the linker attachment to the surface cysteines. GuHCl was used to produce the specific unfolding of the C-terminal helix as described in Chapter 4 to simulate the isomerisation of the linker.

The maleimide attachment was performed at 4 °C and 25 °C to investigate the temperature dependence of the folded state of the C-terminal helix. Excess maleimide was removed using a PD-10 column. A control sample was prepared at the same concentration but was not reacted with maleimide. All 3 samples were partially denatured using the following concentrations of GuHCl: 0, 0.8, 1.6, 2 and 4 M. The samples were incubated on ice for a minimum of 30 minutes before addition of the Ellman's reagent. The UV absorbance was measured for each sample, baselined

Table 5.2: UV absorbance at 412 nm for CTPRy3i11 I97C at different concentrations of GuHCl.

Sample	**GuHCl Concentration**				
	0 M	0.8 M	1.6 M	2 M	4 M
No maleimide	0.289	0.276	0.282	0.291	0.308
4 °C	-0.005	-0.001	0.002	0.016	0.013
25 °C	-0.001	-0.001	0.001	0.005	0.005

against an equivalent concentration of GuHCl.

5.4.5.0.4 Results and Discussion

The absorbances for both the 4 °C and the 25 °C samples were identical at all concentrations of GuHCl. This indicates that the maleimide has reacted with the I97 cysteine, and is solvent accessible. In the control sample there was a small increase (approximately 6 %) in the absorbance between the 0 M and the 4 M samples. At 4 M GuHCl the protein will be fully denatured, thus there should be a 50 % increase in the absorbance observed. As this was not the result, we conclude that the cysteine at position 97 is solvent accessible. Although the I97C mutation was not suitable as a detection method for the unfolding of the C-terminal helix, there are other residues that could be suitable.

5.4.6 Conclusions

UV absorption spectra show that the AB-photolinker does undergo a trans→cis isomerisation, when attached to the C-terminal helix of a CTPR. We were unable to observe a significant change in the helical content of the protein via NMR. We only observed a small change in the CD spectra. Other studies using similar linkers reported only small changes in helicity. In one paper, a 50 % decrease in CD signal was observed for a single α-helical peptide, where the whole structure was being disrupted by a photolinker [123]. In the system described here, full unfolding of

the C-terminal helix would produce a 14 % decrease in signal (assuming all the AB-photolinker isomerises). We observed a 7 % decrease in CD signal, therefore we believe folded structure of the helix is being distorted.

Chapter 6

Conclusions and Further Work

In this thesis I have shown how the CTPR scaffold can be engineered with a view to creating biomaterials. In Chapter 3, I described the design of a system for creating self-assembling fibres from CTPR3ΔS monomers using Native Chemical Ligation. In Chapter 4, the unfolding characteristics of a series of deletion mutants was studied and this data fit to a Heteropolymer Ising model. This led to the successful design and implementation of a thermodynamic switch into the CTPR2 and CTPR3 proteins. Finally, in Chapter 5, we explored the use of a photoisomerisable linker to unfold the C-terminal helix of CTPR3.

6.1 Self-Assembly of CTPR Modules

The aim of this chapter was to create a self-assembling system using recombinant CTPR units. We have demonstrated that we can make fibres from CTPR monomers that have been covalently bonded together using Native Chemical Ligation. Using this reaction we were able to create structures ranging in size from short oligomers to fibres up to microns in length. Our system is the first example of NCL being used to link repeat protein monomers into fibres. We used a modified pTWIN expression vector to produce a monomer (His-Xa-CTPR3ΔS-thioester) that undergoes NCL

when triggered by the Factor Xa cleavage of the N-terminal His-tag. The protein successfully oligomerised into long oligomers. We followed the reaction using SDS-PAGE, size exclusion chromatography, MS, CD and TEM. Using TEM we were able to observe fibres that were formed as a result of the oligomerisation.

We found that the oligomerisation was sensitive to the reaction conditions. The addition of NaCl to the reaction buffer resulted in an increase in the rate of reaction. This is potentially because the addition of salt provides a more favourable environment for the initiation step (Factor Xa cleavage). After observing many polymerisation reactions, it has become clear that the initiation step is the rate limiting step. This process is limited by the concentration of Factor Xa. Whilst this effect could be overcome by increasing the concentration of Factor Xa, this would vastly increase the cost of making the fibres. Instead of using Factor Xa to initiate the reaction, we could redesign the protein to use a modified thrombin or TEV site that still leaves the required cysteine at the N-terminus. As we observed that the Factor Xa cleaved at an alternative site to the designed one, changing the enzyme to thrombin or TEV would eliminate both the problem of cost and improper cleavage.

In addition to increasing the yield of fibres, there are two other developments that could be made: adding functionality to the fibres and controlling the polymerisation process. Functionality could be created by engineering peptide binding sites into the CTPR monomers. Currently, work being conducted by Joseph Harvey is exploring the development of a system where there is controlled addition of the monomeric units, in a similar way to solid state peptide synthesis. This would allow the length of the fibres to be controlled and limited to a specific number. Individual control over each monomer in the sequence would make it possible to create heterogeneous fibres, containing CTPRs with different binding motifs.

6.2 Introducing a Thermodynamic Switch into CTPR2 and CTPR3

We investigated the chemical denaturation data for a series of deletion mutants for both the CTPRan and the CTPRn series. This data was then fit to a Heteropolymer Ising model. The model was able to calculate the change in free energy for 3 types of helices in the protein: N-terminal, C-terminal and internal. With this data it was possible to simulate the unfolding curve of different combinations of helices. We were also able to calculate the population of partially unfolded species.

The population of protein with the C-terminal helix unfolded but internal helices folded (S-switch population) was calculated for each full length CTPR. The Heteropolymer Ising model was then used to calculate the S-switch population for a chimera of the 2 CTPR series, i.e. a protein with the internal helices of CTPRn and the C-terminal helix of CTPRan. The chimera of CTPR3 and CTPRa3 involves the mutation of the final loop before the C-terminal helix from `PNN` to `PRS`. This protein was termed CTPRy3 and was predicted to have an S-switch population of $\approx$ 50 % at 2.8 M GuHCl, with only $\approx$ 15 % of further helical unfolding. The chemical denaturation curve of CTPRy3 corresponded well to the simulated curve showing that the model is able to make accurate predictions. CTPRy3 was evaluated using 2D NMR and this showed that there was a conformational change in the C-terminal helix at 1.6 M GuHCl. The same mutation was placed in CTPR2 to create CTPRy2. Whilst both CTPRy3 and CTPRy2 displayed a conformational change at a specific concentration of denaturant, the switch in CTPRy3 was more defined. As a result of identifying and characterising the effect of the `PNN` to `PRS` mutation, we have incorporated it into the design of two of the AB-photolinker proteins.

6.3 Photoisomerisable Azobenzene Crosslinker

Here we investigated using a photoisomerisable linker to disrupt the folding of the C-terminal helix of CTPR3. Four double cysteine mutants of CTPR3 were designed. A photoisomerisable linker was ligated onto the cysteine residues located in the C-terminal helix. The linker reversibly isomerised when exposed to UV light at 365 nm. The change in UV absorbance at 365 nm confirmed that the linker was isomerising. We used CD to observe the structural changes as a result of the isomerisation. We saw a small change in the ellipticity at 222 nm indicating that there is some disruption of the folding of the C-terminal helix. We also employed 2D NMR to further probe the structural changes occurring. However, we were unable to observe any significant change in the HSQC spectra. This would have confirmed a structural change in the C-terminal helix. We explored the possibility of masking a cysteine under the C-terminal helix that, after isomerisation, is solvent exposed. However, the cysteine mutation, I97C, was accessible to the solvent at all times. Further investigations could identify a more suitable location for the masked cysteine.

6.4 Future Directions

6.4.1 Genetically-encoded Click Chemistry

We have demonstrated that it is possible to polymerise CTPRs by forming covalent bonds between monomers following enzymatic cleavage. In chapter 4 we successfully designed a thermodynamic switch into CTPR3 and CTPR2. The switch was designed to reveal a dimerisation interface, however no oligomerisation was observed. For polymerisation of CTPRs to take place, a stronger driving force is required than simply the interaction between interfaces at the N and C-termini. We would believe that we can combine these two processes: partial unfolding and covalent

oligomerisation to create a system that partially unfolds, reacts and subsequently oligomerises. One method we propose involves the use of our switching protein, CT-PRy3, in conjunction with copper-free click chemistry to form the covalent linkages between the monomers, Figure 6.1

A copper-free click reaction involves a cyclo-addition reaction between an azide and a strained ring system, di-benzocyclooctyne (DBCO). The reaction is highly specific, produces no harmful by-products and can be performed in aqueous solvents. Previously the azide containing groups had to be attached to proteins through ligation to a side chain of an amino acid. However, it is now possible to incorporate azide-containing non-natural amino acids (NNAA) at specific sites in proteins through the use of orthogonal t-RNA synthetases. The process of incorporating a NNAA involves placing an amber stop codon in the gene at the site of the desired mutation. The gene is recombinantly expressed in the presence of a designed t-RNA synthetase and the NNAA. When the amber stop codon is read on the bacterial ribosome, the NNAA is added by the t-RNA synthetase. The DBCO functionality is incorporated through the reactive groups of cysteine or lysine residues, using maleimide ot N-hydroxysuccinamide respectively.

The proposed monomer would possess both click reactants. At the N-terminus, a DBCO group will be ligated to an N-terminal cysteine. The C-terminal repeat will contain an azide, that is prevented from reaction by the C-terminal helix. In the reaction scheme, firstly the C-terminal helix unfolds to expose the reactive azide, Figure 6.1B. A click reaction can then occur between the azide and the DBCO, to create a dimer that is covalently linked together.

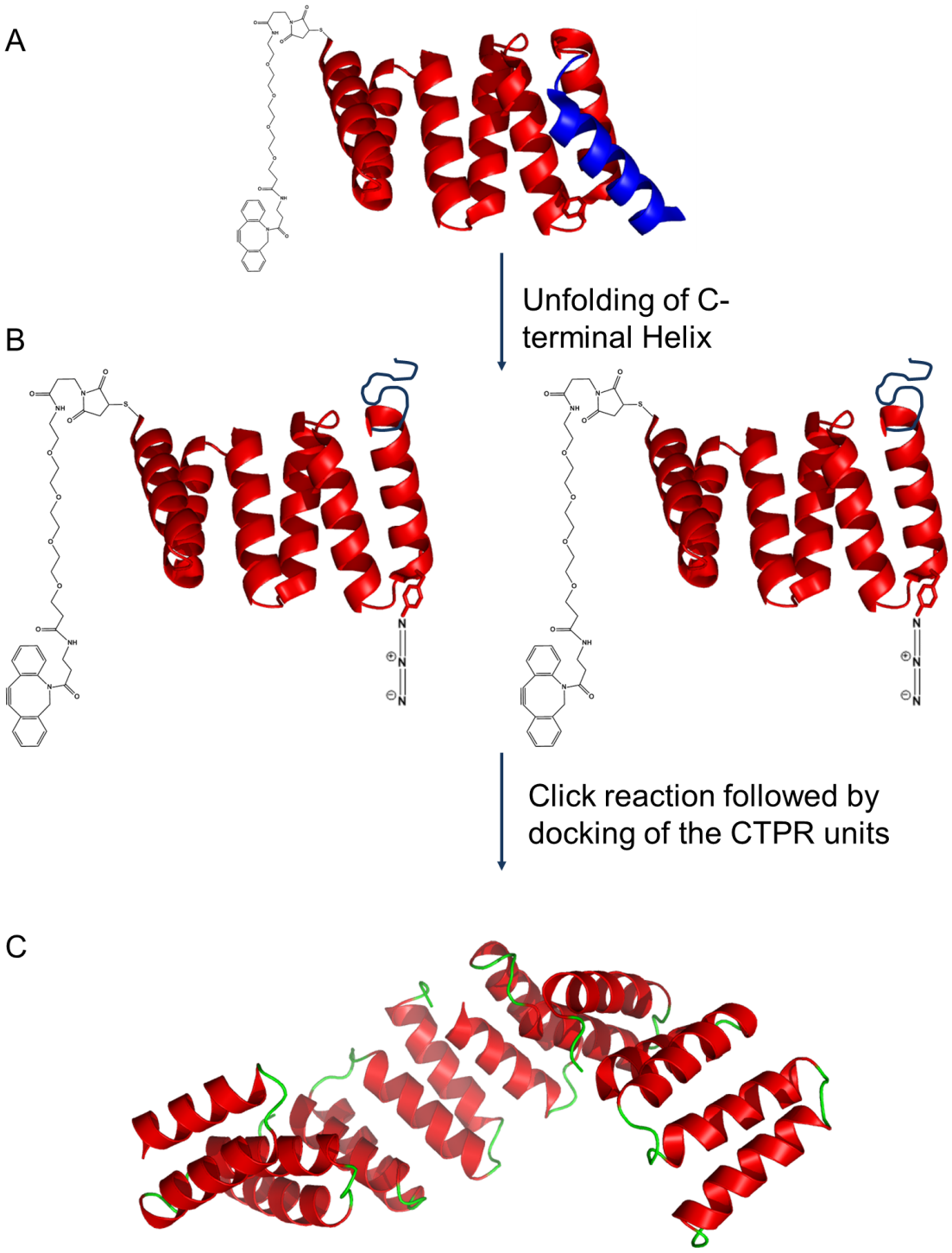

Figure 6.1: (A) Schematic representation of a proposed click chemistry ready monomer (constructed using 1NA0). A DBCO group is ligated onto a cysteine residue at the N-terminus and para azidophenylalanine is protected from reaction by the C-terminal helix shown in blue. (B) When exposed to heat or denaturant the C-terminal helix is unfolded to expose the azide group to reaction with the DBCO group on another monomer. (C) The click reaction locks the units together and they can dock together though compatible interfaces to form an ordered polymer in a specific head-to-tail manner.

6.4.2 Protein Nanocages

We have shown that we can make fibres using CTPRs in a self-assembling system. Now the focus of interest is in combining the CTPR scaffold with other self-assembling proteins to extend the tool kit of monomeric building blocks. This will allow a more diverse range of structures to be formed. The CTPR scaffold could be used in the design and assembly of protein nano-cages. A protein nano-cage is comprised of self-assembling units that associate to form an encapsulation with dimensions in the nanometre range. A two-component protein nano-cage contains one type of protein that forms the vertices and another that forms the edges or sides of the cage. In work that is currently ongoing, a CTPR3 unit will be used to form the side of a cage. CTPRs are a good choice as the sides of a cage as they are structurally rigid and the N and C-terminus can easily be functionalised without disrupting the fold. The vertices of the proposed nano-cage involve the homo-trimer, Monofoil-4-P (M4P) comprised of three 42 amino acid peptide subunits that spontaneously associate [124]. The ability to associate will be exploited in the self-assembly process. Like the CTPRs, the M4P has free termini that can be used to connect the M4P to other proteins and peptides.

The designed CTPR3-M4P nano-cage involves the C-terminus of the CTPR3 unit fused to the M4P, Figure 6.2. When the M4P associates, an open half-cage is formed. By adding functionality at the N-terminus of the CTPR3 it will be possible to link the units together and close the cage.

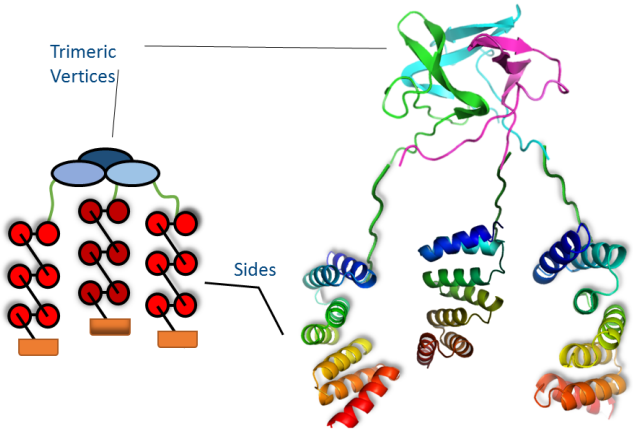

Figure 6.2: Schematic of the formation of a CTPR3-M4P trimer (constructed from 1NA0 and 3OL0). The Monofoil-4-P forms a trimer with the CTPR3 units linked via the C-terminus. By engineering functionality to the N-terminus of the CTPR3 it would be possible to form 'closed' cages. Figure produced by James Wright.

Appendix

7.0.3 Primers

7.0.3.1 CTPRy2 and CTPRy3

Primer	
Forward	5’ CTTGAGCTCGATCCAAgatcTGCTGAAGCTAAACAG 3’
Reverse	5’ CTGTTTAGCTTCAGCAgatcTTGGATCGAGCTCAAG 3’

7.0.3.2 CTPR2ΔS and CTPR3ΔS

Primer	
Forward	5’GAGCTCGATCCAAATAATtgaGAAGCTAAACAGAACCTG 3’
Reverse	5’CAGGTTCTGTTTAGCTTCTCAATTATTTGGATCGAGCTC 3’

7.0.3.3 Minus A Helix deletion mutants of CTPR2 and CTPR3

Primer	
Forward	5’-TGCGAATACTACCAGAAAGCGCTGGAACTG-3’
Reverse	5’-CGCTTCGTCGTAGTCCCCCTG-3’

7.0.3.4 Cysteine mutants of Photolinker construct i11PRS

Primer	
Forward	5'-TGCGAATACTACCAGAAAGCGCTGGAACTG-3'
Reverse	5'-CGCTTCGTCGTAGTCCCCCTG-3'

7.0.4 Preparing Agarose gels

A 1 % Agarose gel was prepared by superheating 1g of Agarose in 100 ml of Tris-Borate-EDTA (TBE) buffer (45 mM Tris Base, 45 mM Boric acid and 10 mM EDTA at pH 8.3) until fully dissolved. Ethidium Bromide was added to a final concentration of 1 μg/ml.

7.0.4.1 Gel preparation

Gels were all 100m x 100mm in size, with either 12 or 16 sample wells. 18% Gels were the most commonly used, the volumes required to make 2 gels are shown in table 7.1

Table 7.1: Volumes required to make 2 18% gels

Componant	**Volume Resolving Gel**	**Volume Stacking Gel**
Tris pH 8.8	5 ml	N/A
Tris pH 6.8	N/A	0.75 ml
Acrylamide	12 ml	1 ml
ddH_2O	2.6 ml	4 ml
10 % SDS	0.2 ml	60 μl
10 % APS	0.2 ml	60 μl
TEMED	8 μl	6 μl

7.0.4.2 Buffers

The buffers used in preparing the gels were: 1.5 M Tris pH 8.8, 1 M Tris pH 6.8, 10 % SDS (w/v), 10 % Ammonium Persulfate (w/v), 30 % Acryamide, The Running

Buffer was prepared using: 144 g Glycine; 30 g Tris Base; 10 g SDS. Dissolved in 1 L ddH_2O and corrected to pH 8.3. Loading buffer was a 2X concentrated stock, containing: 100 mM Tris pH 6.8; 200 mM DTT; 4 % SDS (w/v); 20 % Glycerol (w/v) and 0.2 % Bromophenol blue (w/v).

7.0.5 Gel of CTPR proteins

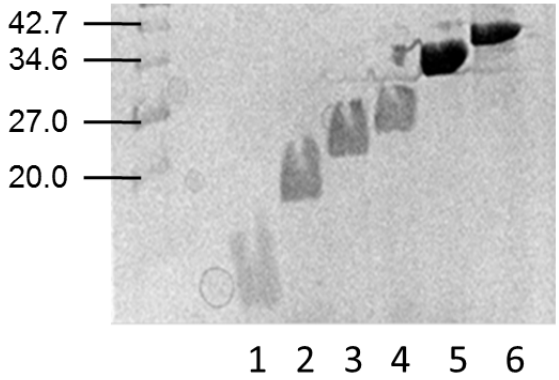

Figure 7.1: 18% SDS-PAGE gel showing lane 1;CTPRa3, lane 2;CTPRa4, lane 3;CTPRa5, lane 4;CTPRa6, lane 5;CTPRa8 and lane 6;CTPRa10. We can see that as the proteins increase in length there is a smaller change in the size observed on the gel.

Bibliography

[1] JM Berg, JL Tymoczko, and L Stryer. *Biochemistry: 5th edition.* W H Freeman, 2002.

[2] William K Purves. *Life, the Science of Biology.* W H Freeman, 2001.

[3] MJ Romao, J Knablein, R Huber, and JJG Moura. Structure and function of molybdopterin containing enzymes. *Progress in Biophysics & Molecular Biology*, 68(2-3):121–144, 1997.

[4] G Layer, DW Heinz, D Jahn, and WD Schubert. Structure and function of radical SAM enzymes. *Current Opinion in Chemical Biology*, 8(5):468–476, 2004.

[5] Loredana Lo Conte, Bart Ailey, Tim JP Hubbard, Steven E Brenner, Alexey G Murzin, and Cyrus Chothia. SCOP: a Structural Classification of Proteins database. *Nucleic Acids Res.*, 28:257–259, 2000.

[6] Andreas D Baxevanis. *The Importance of Biological Databases in Biological Discovery.* John Wiley Sons, Inc., 2002.

[7] Brandon C and Tooze J. *Introduction to Protein Structure.* Garland, 1999.

[8] Ewan RG Main, Yong Xiong, Melanie J Cocco, Luca D'Andrea, and Lynne Regan. Design of Stable alpha-Helical Arrays from an Idealized TPR Motif . *Structure*, 11(5):497 – 508, 2003.

[9] Yalda Javadi and Ewan R G Main. Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins. *PNAS*, 106(41):17383–17388, 2009.

[10] Miguel A Andrade, Carolina Perez-Iratxeta, and Chris P Ponting. Protein repeats: Structures, functions, and evolution. *Journal of Structural Biology*, 134(2–3):117 – 131, 2001.

[11] Ellen Kloss, Naomi Courtemanche, and Doug Barrick. Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. *Archives of Biochemistry and Biophysics*, 469(1):83 – 99, 2008.

[12] Yalda Javadi and Laura S Itzhaki. Tandem-repeat proteins: regularity plus modularity equals design-ability. *Current Opinion in Structural Biology*, 23(4):622 – 631, 2013.

[13] Ewan RG Main, Sophie E Jackson, and Lynne Regan. The folding and design of repeat proteins: reaching a consensus. *Current Opinion in Structural Biology*, 13(4):482 – 489, 2003.

[14] Jungyong Nam, Won Mah, and Eunjoon Kim. The SALM/Lrfn family of leucine-rich repeat-containing cell adhesion molecules. *Seminars in Cell Developmental Biology*, 22(5):492 – 498, 2011.

[15] Peer RE Mittl and Wulf Schneider-Brachert. Sel1-like repeat proteins in signal transduction. *Cellular Signalling*, 19(1):20 – 31, 2007.

[16] Juqiang Yan, Jing Wang, and Hong Zhang. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *The Plant Journal*, 29(2):193–202, 2002.

[17] V Shanmugam. Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms. *Microbiological Research*, 160(1):83 – 94, 2005.

[18] Brody J DeYoung, Dong Qi, Sang-Hee Kim, Thomas P Burke, and Roger W Innes. Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cellular Microbiology*, 14(7):1071–1084, 2012.

[19] Leila K Mosavi, Tobin J Cammett, Daniel C Desrosiers, and Zheng-yu Peng. The ankyrin repeat as molecular architecture for protein recognition. *Protein Science*, 13(6):1435–1448, 2004.

[20] Steven G Sedgwick and Stephen J Smerdon. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in Biochemical Sciences*, 24(8):311 – 316, 1999.

[21] S Gorina and NP Pavletich. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science*, 274(5289):1001–1005, 1996.

[22] Purevjav Enkhbayar, Masakatsu Kamiya, Mitsuru Osaki, Takeshi Matsumoto, and Norio Matsushima. Structural principles of leucine-rich repeat (LRR) proteins. *Proteins: Structure, Function, and Bioinformatics*, 54(3):394–403, 2004.

[23] DA Jones and JDG Jones. The role of leucine-rich repeat proteins in plant defences. volume 24 of *Advances in Botanical Research*, pages 89 – 167. Academic Press, 1997.

[24] RC Hillig, L Renault, IR Vetter, T Drell, A Wittinghofer, and J Becker. The crystal structure of rna1p: A new fold for a GTPase-activating protein. *Molecular Cell*, 3(6):781–791, 1999.

[25] Luca D D'Andrea and Lynne Regan. TPR proteins: the versatile helix . *Trends in Biochemical Sciences*, 28(12):655 – 662, 2003.

[26] C Scheufler, A Brinker, G Bourenkov, S Pegoraro, L Moroder, H Bartunik, FU Hartl, and I Moarefi. Structure of TPR domain-peptide complexes: Crit-

ical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell*, 101(2):199–210, 2000.

[27] Schedl P Riggleman B, Wieschaus E. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a drosophila segment polarity gene. *Genes Dev.*, 3:96–113, 1989.

[28] Miguel A Andrade and Peer Bork. HEAT repeats in the Huntington's disease protein. *Nat. Genet.*, 11:115–116, 1995.

[29] G Cingolani, C Petosa, K Weis, and CW Muller. Structure of importin-beta bound to tbe IBB domain of importin-alpha. *Nature*, 399(6733):221–229, 1999.

[30] Adam J Bogdanove and Daniel F Voytas. TAL Effectors: Customizable Proteins for DNA Targeting. *Science*, 333(6051):1843–1846, 2011.

[31] Heidi Scholze and Jens Boch. TAL effectors are remote controls for gene activation . *Current Opinion in Microbiology*, 14(1):47 – 53, 2011.

[32] Amanda Nga-Sze Mak, Philip Bradley, Adam J Bogdanove, and Barry L Stoddard. TAL effectors: function, structure, engineering and applications. *Current Opinion in Structural Biology*, 23(1):93 – 99, 2013.

[33] Asuka Eguchi, Garrett O Lee, Fang Wan, Graham S Erwin, and Aseem Z Ansari. Controlling gene networks and cell fate with precision-targeted DNA-binding proteins and small-molecule-based genome readers. *Biochemical Journal*, 462(3):397–413, 2014.

[34] John Jenkins, Olga Mayans, and Richard Pickersgill. Structure and Evolution of Parallel beta-Helix Proteins . *Journal of Structural Biology*, 122(1–2):236 – 246, 1998.

[35] John Jenkins and Richard Pickersgill. The architecture of parallel beta-helices and related folds. *Progress in Biophysics and Molecular Biology*, 77(2):111 –

175, 2001.

[36] Shuaiqi Guo, Christopher P Garnham, John C Whitney, Laurie A Graham, and Peter L Davies. Re-Evaluation of a Bacterial Antifreeze Protein as an Adhesin with Ice-Binding Activity. *PLoS ONE*, 7(11):48805, 2012.

[37] Yih-Cherng Liou, Ante Tocilj, Peter L Davies, and Zongchao Jia. Mimicry of ice structure by surface hydroxyls and water of a β-helix antifreeze protein. *Letter to Nature*, 406:320–322, 2000.

[38] H Kaspar Binz, Michael T Stumpp, Patrik Forrer, Patrick Amstutz, and Andreas Plückthun. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology*, 332(2):489 – 503, 2003.

[39] LK Mosavi, DL Minor, and ZY Peng. Consensus-derived structural determinants of the ankyrin repeat motif. *PNAS*, 99(25):16029–16034, 2002.

[40] Aitziber L Cortajarena, Jimin Wang, and Lynne Regan. Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand. *FEBS Journal*, 277(4):1058–1066, 2010.

[41] Andreas Plückthun. Ribosome display: A perspective. In Julie A. Douthwaite and Ronald H. Jackson, editors, *Ribosome Display and Related Technologies*, volume 805 of *Methods in Molecular Biology*, pages 3–28. Springer New York, 2012.

[42] Christian Zahnd, Emanuel Wyler, Jochen M Schwenk, Daniel Steiner, Michael C Lawrence, Neil M McKern, Frédéric Pecorari, Colin W Ward, Thomas O Joos, and Andreas Plückthun. A Designed Ankyrin Repeat Protein Evolved to Picomolar Affinity to Her2 . *Journal of Molecular Biology*, 369(4):1015 – 1028, 2007.

[43] Lutz Kummer, Petra Parizek, Peter Rube, Bastian Millgramm, Anke Prinz, Peer RE Mittl, Melanie Kaufholz, Bastian Zimmermann, Friedrich W Herberg, and Andreas Plueckthun. Structural and functional analysis of phosphorylation-specific binders of the kinase ERK from designed ankyrin repeat protein libraries. *PNAS*, 109(34):2248–2257, 2012.

[44] Johannes Schilling, Jendrik Schoeppe, and Andreas Plueckthun. From DARPins to LoopDARPins: Novel LoopDARPin Design Allows the Selection of Low Picomolar Binders in a Single Round of Ribosome Display. *Journal of Molecular Biology*, 426(3):691–721, 2014.

[45] Romer, L and Scheibel, T. The elaborate structure of spider silk: Structure and function of a natural high performance fiber. *Prion*, 2:154–161, 2008.

[46] Derek N Woolfson. Building fibrous biomaterials from alpha-helical and collagen-like coiled-coil peptides. *Peptide Science*, 94(1):118–127, 2010.

[47] Thomas Scheibel. Protein fibers as performance proteins: new technologies and applications . *Current Opinion in Biotechnology*, 16(4):427 – 433, 2005. Protein technologies and commercial enzymes.

[48] Honggang Cui, Matthew J Webber, and Samuel I Stupp. Self-assembly of peptide amphiphiles: From molecules to nanostructures to biomaterials. *Peptide Science*, 94(1):1–18, 2010.

[49] Der, Bryan S and Kuhlman, Brian. Cages from Coils. *Nature Biotechnology*, pages 809–810, 2013.

[50] Nasim Annabi, Suzanne M. Mithieux, Gulden Camci-Unal, Mehmet R. Dokmeci, Anthony S. Weiss, and Ali Khademhosseini. Elastomeric recombinant protein-based biomaterials. *Biochemical Engineering Journal*, 77(0):110 – 118, 2013.

[51] Suzanne M. Mithieux, John E.J. Rasko, and Anthony S. Weiss. Synthetic elastin hydrogels derived from massive elastic assemblies of self-organized human protein monomers. *Biomaterials*, 25(20):4921 – 4927, 2004. Focus on Biomaterials Science in Australia.

[52] Nobuyuki Higashi, Kaori Yasufuku, Yusaku Matsuo, Takahiro Matsumoto, and Tomoyuki Koga. Thermo-responsive multilayer films from ionic polymers with elastin-like peptide as graft chains. *Colloids and Interface Science Communications*, 1(0):50 – 53, 2014.

[53] Robert J. Mart, Rachel D. Osborne, Molly M. Stevens, and Rein V. Ulijn. Peptide-based stimuli-responsive biomaterials. *SOFT MATTER*, 2(10):822–835, OCT 7 2006.

[54] Jindrich Kopecek. Hydrogel biomaterials: A smart future? *BIOMATERIALS*, 28(34):5185–5192, DEC 2007.

[55] Tijana Z. Grove, Chinedum O Osuji, Jason D Forster, Eric R Dufresne, and Lynne Regan. Stimuli-responsive smart gels realized via modular protein design. *Journal of the American Chemical Society*, 132(40):14024–14026, 2010.

[56] Derek N Woolfson and Zahra N Mahmoud. More than just bare scaffolds: towards multi-component and decorated fibrous biomaterials. *Chem. Soc. Rev.*, 39:3464–3479, 2010.

[57] Kyle Morris and Louise Serpell. From natural to designer self-assembling biopolymers the structural characterisation of fibrous proteins and peptides using fibre diffraction. *Chem. Soc. Rev.*, 39:3445–3453, 2010.

[58] Cooper GM. *The Cell: A Molecular Approach*. Sinauer Associates, 2000.

[59] CJ Tsai, J Zheng, D Zanuy, N Haspel, and R Nussinov. Principles of nanostructure design with protein building blocks. *Proteins*, 68, 2007.

[60] Charles J Bowerman and Bradley L Nilsson. Review self-assembly of amphipathic beta-sheet peptides: Insights and applications. *Peptide Science*, 98(3):169–184, 2012.

[61] Derek N Woolfson, Gail J Bartlett, Marc Bruning, and Andrew R Thomson. New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. *Current Opinion in Structural Biology*, 22(4):432 – 441, 2012. Engineering and design / Membranes.

[62] Silvia Cavalli, Fernando Albericio, and Alexander Kros. Amphiphilic peptides and their cross-disciplinary role as building blocks for nanoscience. *Chem. Soc. Rev.*, 39:241–263, 2010.

[63] Xiao-Ming Zhou, Aiman Entwistle, Hong Zhang, Antony P Jackson, Thomas O Mason, Ulyana Shimanovich, Tuomas P J Knowles, Andrew T Smith, Elizabeth B Sawyer, and Sarah Perrett. Self-Assembly of Amyloid Fibrils That Display Active Enzymes. *Chem. Cat. Chem.*, 6(7):1961–1968, 2014.

[64] Seulgi Kim, Song Yi Bae, Bun Yeoul Lee, and T. Doohun Kim. Coaggregation of amyloid fibrils for the preparation of stable and immobilized enzymes . *Analytical Biochemistry*, 421(2):776 – 778, 2012.

[65] T Luhrs, C Ritter, M Adrian, D Riek-Loher, B Bohrmann, H Doeli, D Schubert, and R Riek. 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *PNAS*, 102(48):17342–17347, 2005.

[66] Jiyuan Yang, Chunyu Xu, Chun Wang, and Jindřich Kopeček. Refolding Hydrogels Self-Assembled from N-(2-Hydroxypropyl)methacrylamide Graft Copolymers by Antiparallel Coiled-Coil Formation. *Biomacromolecules*, 7(4):1187–1195, 2006.

[67] David Papapostolou, Andrew M Smith, Edward DT Atkins, Seb J Oliver,

Maxim G Ryadnov, Louise C Serpell, and Derek N Woolfson. Engineering nanoscale order into a designed protein fiber. *PNAS*, 104(26):10853–10858, 2007.

[68] Mehrban, N and Zhu, B and Tamagnini, F and Young, F I and Wasmuth, A and Hudson, K L and Thomson, A R and Birchall, M A and Randall, A D and Song, B and Woolfson, D N. Functionalized Alpha-helical Peptide Hydrogels for Neural Tissue Engineering. *ACS Biomaterials Science Engineering*, 2015.

[69] Jordan M Fletcher, Robert L Harniman, Frederick RH Barnes, Aimee L Boyle, Andrew Collins, Judith Mantell, Thomas H Sharp, Massimo Antognozzi, Paula J Booth, Noah Linden, Mervyn J Miles, Richard B Sessions, Paul Verkade, and Derek N Woolfson. Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science*, 340(6132):595–599, 2013.

[70] JD Hartgerink, E Beniash, and SI Stupp. Peptide-amphiphile nanofibers: A versatile scaffold for the preparation of self-assembling materials. *PNAS*, 99(8):5133–5138, 2002.

[71] Timothy D Sargeant, Mustafa O Guler, Scott M Oppenheimer, Alvaro Mata, Robert L Satcher, David C Dunand, and Samuel I Stupp. Hybrid bone implants: Self-assembly of peptide amphiphile nanofibers within porous titanium. *Biomaterials*, 29(2):161–171, 2008.

[72] Aitziber L Cortajarena and Lynne Regan. Calorimetric study of a series of designed repeat proteins: Modular structure and modular folding. *Protein Science*, 20(2):336–340, 2011.

[73] Tural Aksel, Ananya Majumdar, and Doug Barrick. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure*, 19(3):349 – 360, 2011.

[74] Randall P Watson, Martin T Christen, Christina Ewald, Fabian Bumbak,

Christian Reichen, Maja Mihajlovic, Elena Schmidt, Peter Guentert, Amedeo Caflisch, Andreas Plueckthun, and Oliver Zerbe. Spontaneous Self-Assembly of Engineered Armadillo Repeat Protein Fragments into a Folded Structure. *Structure*, 22(7):985–995, 2014.

[75] Ewan R G Main, Katherine Stott, Sophie E Jackson, and Lynne Regan. Local and long-range stability in tandemly arrayed tetratricopeptide repeats. *PNAS*, 102(16):5721–5726, 2005.

[76] Laura S Itzhaki and Alan R Lowe. From artificial antibodies to nanosprings. In JacquelineM. Matthews, editor, *Protein Dimerization and Oligomerization in Biology*, volume 747 of *Advances in Experimental Medicine and Biology*, pages 153–166. Springer New York, 2012.

[77] Laura S Lowe, Alan R band Itzhaki. Rational redesign of the folding pathway of a modular protein. *PNAS*, 104(8):2679–2684, 2007.

[78] J J Phillips, Y Javadi, C Millership, and E R G Main. Modulation of the multistate folding of designed tpr proteins through intrinsic and extrinsic factors. *Protein Science*, 21(3):327–338, 2012.

[79] Nicolas D Werbeck and Laura S Itzhaki. Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. *PNAS*, 104(19):7863–7868, 2007.

[80] Alexandre Chenal, J Iñaki Guijarro, Bertrand Raynal, Muriel Delepierre, and Daniel Ladant. Rtx calcium binding motifs are intrinsically disordered in the absence of calcium: Implication for protein secretion. *Journal of Biological Chemistry*, 284(3):1781–1789, 2009.

[81] Kevin Dooley, Yang Hee Kim, Hoang D Lu, Raymond Tu, and Scott Banta. Engineering of an environmentally responsive beta roll peptide for use as

a calcium-dependent cross-linking domain for peptide hydrogel formation. *Biomacromolecules*, 13(6):1758–1764, 2012.

[82] Andrew J Scotter, Meng Guo, Melanie M Tomczak, Margaret E Daley, Robert L Campbell, Richard J Oko, David A Bateman, Avijit Chakrabartty, Brian D Sykes, and Peter L Davies. Metal ion-dependent, reversible, protein filament formation by designed beta-roll polypeptides. *BMC Structural Biology*, 7(1), 2007.

[83] Tijana Z Grove, Lynne Regan, and Aitziber L Cortajarena. Nanostructured functional films from engineered repeat proteins. *Journal of The Royal Society Interface*, 10(83), 2013.

[84] ChristianPR Hackenberger and Dirk Schwarzer. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angewandte Chemie International Edition*, 47(52):10030–10074, 2008.

[85] Richard Ramsden, Luther Arms, Trisha Davis, and Eric Muller. An intein with genetically selectable markers provides a new approach to internally label proteins with GFP. *BMC Biotechnology*, 11(1):71, 2011.

[86] Robert M Hagan, Ragnar Björnsson, Stephen A McMahon, Benjamin Schomburg, Vickie Braithwaite, Michael Bühl, James H Naismith, and Ulrich Schwarz-Linek. Nmr spectroscopic and theoretical analysis of a spontaneously formed lys–asp isopeptide bond. *Angewandte Chemie International Edition*, 49(45):8421–8425, 2010.

[87] Bijan Zakeri, Jacob O Fierer, Emrah Celik, Emily C Chittock, Ulrich Schwarz-Linek, Vincent T Moy, and Mark Howarth. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *PNAS*, 2012.

[88] Jacob O Fierer, Gianluca Veggiani, and Mark Howarth. Spyligase pep-

tide–peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *PNAS*, 111(13):1176–1181, 2014.

[89] Michael Fairhead, Gianluca Veggiani, Melissa Lever, Jun Yan, Dejan Mesner, Carol V Robinson, Omer Dushek, P Anton van der Merwe, and Mark Howarth. Spyavidin hubs enable precise and ultrastable orthogonal nanoassembly. *Journal of the American Chemical Society*, 136(35):12355–12363, 2014.

[90] Ryo Matsunaga, Saeko Yanaka, Satoru Nagatoishi, and Kouhei Tsumoto. Hyperthin nanochains composed of self-polymerizing protein shackles. *Nature Communications*, 4, 2013.

[91] HC Kolb, MG Finn, and KB Sharpless. Click chemistry: Diverse chemical function from a few good reactions. *Angewandte Chemie International Edition*, 40(11):2004, 2001.

[92] Kido Nwe and Martin W Brechbiel. Growing Applications of "Click Chemistry" for Bioconjugation in Contemporary Biomedical Research. *Cancer Biotherapy and Radiopharmaceuticals*, 24(3):289–302, 2009.

[93] Q Wang, TR Chan, R Hilgraf, VV Fokin, KB Sharpless, and MG Finn. Bioconjugation by copper(I)-catalyzed azide-alkyne [3+2] cycloaddition. *Journal of the American Chemical Society*, 125(11):3192–3193, 2003.

[94] A Deiters, TA Cropp, M Mukherji, JW Chin, JC Anderson, and PG Schultz. Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *Journal of the American Chemical Society*, 125(39):11782–11783, 2003.

[95] Manuel Simon, Uwe Zangemeister-Wittke, and Andreas Plückthun. Facile double-functionalization of designed ankyrin repeat proteins using click and thiol chemistries. *Bioconjugate Chemistry*, 23(2):279–286, 2012.

[96] Invitrogen. *pProEX HTb Expression Vector*. 1996.

[97] Invitrogen. *pTrcHis and pTrcHis2 TOPO TA Expression Kits Five-minute cloning of Taqpolymerase-amplified PCR products for expression in E. coli*, volume J. 2006.

[98] Fermentas. *Thermo Scientific GeneJET Gel Extraction Kit Manual*. 2013.

[99] Qiagen. *QIAprep Miniprep Handbook*, volume 2. 2006.

[100] Fermentas. *Thermo Scientific GeneJET Plasmid Miniprep Kit Manual*. 2013.

[101] Stratagene. *QuikChange Site-Directed Mutagenesis Kit Instruction Manual*.

[102] Gattiker A Duvaud S Wilkins MR Appel RD Bairoch A Gasteiger E, Hoogland C. *The Proteomics Protocols Handbook*. Humana Press, 2005.

[103] C. Tanford. Protein denaturation. *Adv. Protein. Chem.*, 24:1–95, 1968.

[104] Katherine W Tripp and Doug Barrick. Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats. *Journal of Molecular Biology*, 365(4):1187 – 1200, 2007.

[105] Tommi Kajander, Aitziber L. Cortajarena, Ewan R G Main, Simon G J Mochrie, and Lynne Regan. A new folding paradigm for repeat proteins. *Journal of the American Chemical Society*, 127(29):10188–10190, 2005.

[106] A Telenti, M Southworth, F Alcaide, S Daugelat, WR Jacobs, and FB Perler. The Mycobacterium xenopi GyrA protein splicing element: Characterization of a minimal. *Journal of Bacteriology*, 179(20):6378–6382, OCT 1997.

[107] Arianna Rath, Mira Glibowicka, Vincent G Nadeau, Gong Chen, and Charles M Deber. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS*, 106(6):1760–1765, 2009.

[108] Richard J . Jenny, Kenneth G Mann, and Roger L Lundblad. A critical review of the methods for cleavage of fusion proteins with thrombin and factor xa. *Protein Expression and Purification*, 31(1):1 – 11, 2003.

[109] Waugh, David S. An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expression and Purification*, 80(2):283–293, 2011.

[110] Tommi Kajander, Aitziber L. Cortajarena, Simon Mochrie, and Lynne Regan. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallographica Section D*, 63(7):800–811, 2007.

[111] Sergey E. Paramonov, Varun Gauba, and Jeffrey D. Hartgerink. Synthesis of Collagen-like Peptide Polymers by Native Chemical Ligation. *Macromolecules*, 38(18):7555–7561, 2005.

[112] Aitziber L Cortajarena, Simon G J Mochrie, and Lynne Regan. Mapping the Energy Landscape of Repeat Proteins using NMR-detected Hydrogen Exchange . *Journal of Molecular Biology*, 379(3):617 – 626, 2008.

[113] Alan R Fersht and Miriana Petrovich. Reply to campos and muñoz: Why phosphate is a bad buffer for guanidinium chloride titrations. *PNAS*, 110(14):E1244–E1245, 2013.

[114] E. Ising. Report on the theory of ferromagnetism. *Zeitschrift Fur Physik*, 31:253–258, 1925.

[115] Mark P Taylor, Wolfgang Paul, and Kurt Binder. Phase transitions of a single polymer chain: A wang–landau simulation study. *The Journal of Chemical Physics*, 131(11):–, 2009.

[116] Ellen Kloss, Naomi Courtemanche, and Doug Barrick. Repeat-protein folding: New insights into origins of cooperativity, stability, and topology. *Archives of Biochemistry and Biophysics*, 469(1):83 – 99, 2008.

[117] Tural Aksel and Doug Barrick. Chapter 4 analysis of repeatprotein folding using nearestneighbor statistical mechanical models. In Jo M. Holt Michael

L. Johnson and Gary K. Ackers, editors, *Biothermodynamics, Part A*, volume 455 of *Methods in Enzymology*, pages 95 – 125. Academic Press, 2009.

[118] Tural Aksel, Ananya Majumdar, and Doug Barrick. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure*, 19(3):349 – 360, 2011.

[119] Nasrollah Rezaei-Ghaleh, Erika Andreetto, Li-Mei Yan, Aphrodite Kapurniotu, and Markus Zweckstetter. Interaction between Amyloid Beta Peptide and an Aggregation Blocker Peptide Mimicking Islet Amyloid Polypeptide. *PLoS ONE*, 6(5):e20289, 05 2011.

[120] Mike P. Williamson. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73(0):1 – 16, 2013.

[121] HM Dhammika Bandara and Shawn C Burdette. Photoisomerization in different classes of azobenzene. *Chem. Soc. Rev.*, 41:1809–1825, 2012.

[122] Andrew A Beharry and G Andrew Woolley. Azobenzene photoswitches for biomolecules. *Chem. Soc. Rev.*, 40:4422–4437, 2011.

[123] Zhihua Zhang, Darcy C Burns, Janet R Kumita, Oliver S Smart, and G Andrew Woolley. A water-soluble azobenzene cross-linker for photocontrol of peptide conformation. *Bioconjugate Chemistry*, 14(4):824–829, 2003.

[124] Jihun Lee, Sachiko I. Blaber, Vikash K. Dubey, and Michael Blaber. A Polypeptide "Building Block" for the -Trefoil Fold Identified by "Top-Down Symmetric Deconstruction" . *Journal of Molecular Biology*, 407(5):744 – 763, 2011.